



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO EM
MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
PROJETO

PROPOSTA DE SOLUÇÃO DE *BUSINESS*
***INTELLIGENCE* NO CONTEXTO DE**
IMIGRAÇÃO

KHRYSTYNA KRUTS

OUTUBRO - 2020



LISBON
SCHOOL OF
ECONOMICS &
MANAGEMENT
UNIVERSIDADE DE LISBOA

MESTRADO EM
MÉTODOS QUANTITATIVOS PARA A DECISÃO
ECONÓMICA E EMPRESARIAL

TRABALHO FINAL DE MESTRADO
PROJETO

PROPOSTA DE SOLUÇÃO DE *BUSINESS*
***INTELLIGENCE* NO CONTEXTO DE**
IMIGRAÇÃO

KHRYSTYNA KRUTS

ORIENTAÇÃO:

Prof. Doutor CARLOS MANUEL JORGE COSTA

OUTUBRO – 2020

AGRADECIMENTOS

Em primeiro lugar, quero agradecer ao Professor Carlos Costa pela sua excelente orientação neste projeto. A sua dedicação, disponibilidade e apoio foram, sem dúvida, essenciais para a realização do presente projeto.

Devo também um agradecimento aos meus pais, por tudo. Em especial à minha mãe por todo o carinho e força que me deu nesta fase e noutras tantas.

Agradeço, ao ISEG que foi a minha segunda casa durante vários anos. Para além de me abrir portas no mercado de trabalho, deu-me a conhecer pessoas excecionais que hoje tenho a sorte de ter como amigos, Bianca e Francisco. À Bianca, pelo apoio incondicional dia após dia durante a realização deste projeto e ao Francisco por me ajudar e animar sempre que possível.

Quero também deixar um agradecimento a todos os meus amigos que tanto me apoiaram nesta fase, em especial à Belinha por ter estado sempre presente e por me ter motivado a terminar o projeto.

LISTA DE ABREVISTURAS, ACRÓNIMOS E SIGLAS

AMA - Agência para a Modernização Administrativa

BD – Base de Dados

BI – *Business Intelligence*

CRISP-DM – *Cross-Industry Standard Process for Data Mining*

DM – *Data Mining*

INE – Instituto Nacional de Estatística

OIM – Organização Internacional para as Migrações

ONG – Organização Não Governamental

PIB – Produto Interno Bruto

SEF – Serviço de Estrangeiros e Fronteiras

SEFSTAT – Portal de Estatística do Serviço de Estrangeiros e Fronteiras

SI – Sistemas de Informação

SSBI – *Self Service Business Intelligence*

SWOT – *Strengths, Weaknesses, Opportunities and Threats*

UE – União Europeia

RESUMO

Atualmente, o crescente volume de dados dificulta o acesso à informação crucial, existindo a necessidade de análises complexas com recurso a Sistemas de Informação. Em resposta a este desafio, a implementação de soluções de *Business Intelligence* tem sido fulcral na compreensão e processamento de um elevado conjunto de dados, transformando-os em informação útil para o auxílio na tomada de decisão.

A investigação seguiu a metodologia CRISP-DM onde foi desenvolvida uma solução de apoio na tomada de decisão em questões relacionadas com a imigração em Portugal, com recurso à aplicação *Microsoft Power BI*.

Nos últimos anos, Portugal tem registado aumentos significativos no número de população estrangeira residente. São diversos os benefícios que advêm da presença deste conjunto de indivíduos o que torna relevante o seu estudo e o desenvolvimento de ferramentas que permitam análises mais rápidas relativas a esta temática.

A solução foi desenvolvida com base em conceitos abordados no estudo literário e é apresentada em forma de *dashboards* que disponibilizam a informação num formato visualmente apelativo e de fácil compreensão.

O cumprimento dos objetivos definidos resultou numa solução funcional cuja aplicabilidade pode ser ajustada a vários setores.

Palavras-chave: *Business Intelligence*, *CRISP-DM*, *dashboard*, dados, informação, tomada de decisão, *Power BI*, imigração.

ABSTRACT

Nowadays, the growing volume of data makes it difficult to access important information, requiring complex Information Systems analysis in order to get the desired information. As an answer to this challenge, the implementation of Business Intelligence solutions has a huge impact in understanding and processing large data sets, transforming them into knowledge, that can be used in decision-making.

The investigation applied the CRISP-DM methodology and using Microsoft Power BI, the decision-making solution was developed in order to support problems related with immigration in Portugal.

Lately, Portugal has registered significant increases in foreigners residing in the country. The benefits that come from that are innumerable, which makes its study and the development of tools to help with its analysis more poignant than ever.

The solution was developed with basis on concepts from literature review and is presented using dashboards, turning the data into a visually more appealing format, that is easy to read and understand.

By accomplishing the defined goals, a functional solution that can be used in various areas was developed.

Keywords: *Business Intelligence, CRISP-DM, dashboard, data, information, decision making, Power BI, immigration.*

ÍNDICE

AGRADECIMENTOS	i
LISTA DE ABREVISTURAS, ACRÓNIMOS E SIGLAS	ii
RESUMO	iii
ABSTRACT	iv
ÍNDICE.....	v
ÍNDICE DE TABELAS	vii
ÍNDICE DE FIGURAS	vii
1. Introdução.....	1
1.1. Enquadramento	1
1.2. Objetivos de Investigação	1
1.3. Abordagem Metodológica	2
1.4. Estrutura do Projeto	4
2. Revisão da Literatura	5
2.1. Imigração em Portugal.....	5
2.1.1. Tendências e Dinâmicas da Imigração em Portugal	6
2.1.2. Fatores <i>Push-Pull</i>	8
2.1.3. Fatores económicos afetados pela imigração	9
2.2. <i>Business Intelligence</i>	11
2.2.1. Conceito de <i>Business Intelligence</i>	11
2.2.2. Modelação Multidimensional.....	14
2.2.3. <i>Data Visualization</i>	17
2.2.4. <i>Software</i>	19
3. Aplicação da Metodologia CRISP-DM.....	19
3.1. <i>Business Understanding</i>	20
3.2. <i>Data Understanding</i>	21

3.3. <i>Data Preparation</i>	25
3.4. <i>Modeling and Visualization</i>	27
3.5. <i>Evaluation</i>	31
3.6. <i>Deployment</i>	32
4. Resultados	33
5. Conclusões, Limitações e Trabalho Futuro	39
5.1. Conclusões	39
5.2. Limitações e Trabalho Futuro	40
6. Referências Bibliográficas	41
7. Anexos.....	45

ÍNDICE DE TABELAS

Tabela 1: Descrição das variáveis.....	24
Tabela 2: Modelo Relacional.....	29
Tabela 3: Análise SWOT da Solução de BI.....	32

ÍNDICE DE FIGURAS

Figura 1: Fases do modelo CRISP-DM.....	3
Figura 2: O Papel dos Sistemas de BI na tomada de decisão.....	12
Figura 3: Estrutura genérica de um esquema em Estrela.....	15
Figura 4: Estrutura genérica de um esquema em Floco de Neve.....	16
Figura 5: Estrutura genérica de um esquema em Constelação.....	17
Figura 6: Importação, Seleção e Transformação dos dados em <i>Power BI</i>	25
Figura 7: Transformação dos dados para uma estrutura em tabela	26
Figura 8: Eliminação de linhas.....	26
Figura 9: Criação de colunas adicionais	27
Figura 10 (a): Criação manual das relações em <i>Power BI</i>	28
Figura 10 (b): Criação manual das relações em <i>Power BI</i>	28
Figura 11: Modelo Relacional.....	29
Figura 12: Esquematização da Solução Desenvolvida.....	31
Figura 13: Janela Introdutória.....	33
Figura 14: Evolução da População Estrangeira em Portugal (2008-2019).....	34
Figura 15: Evolução da População Estrangeira por Distrito (2008- 2019)	35
Figura 16: Nacionalidades mais representativas em Portugal.....	36

Figura 17: Comparação dos Indicadores Económicos.....	37
Figura 18: Aquisição da Nacionalidade Portuguesa.....	38
Figura 19: Guia para a escolha do gráfico.....	45
Figura 20: <i>Magic Quadrant for Analytics and Business Intelligence Platforms</i>	45
Figura 21: Aplicação dos filtros “Género” - M e “Ano” - 2018.....	46
Figura 22: População estrangeira por distrito e a respetiva percentagem por género.....	46
Figura 23: Comparação da evolução da população estrangeira entre dois distritos.....	47

1. Introdução

1.1. Enquadramento

As migrações constituem uma das mais relevantes influências sobre a transformação social em todo o mundo (Góis & Marques, 2018). Os fluxos migratórios afetam diretamente a economia de um país tendo um impacto notório nas suas taxas de crescimento, nas relações comerciais e balanças de pagamentos, na procura de educação, dos serviços de saúde e infraestruturas sociais. Por outro lado, as condições políticas, sociais e culturais também são altamente afetadas (Cholezas & Tsakloglou, 2009).

Em Portugal, os benefícios da imigração são observados em diversos âmbitos. Desempenha um papel importante na atenuação do decréscimo populacional e no rejuvenescimento das estruturas etárias (Góis, 2019). A nível económico, o impacto da imigração no comércio português é, também, positivo (Faustino & Peixoto, 2009). Por outro lado, a presença de pessoas com diferentes origens promove o progresso cultural do país (Góis, 2019).

O impacto que os fluxos migratórios têm no país de destino torna relevante o seu estudo, designadamente no caso português (Rodrigues & Ferreira, 2014). Desta forma, é benéfico desenvolver ferramentas que auxiliem a tomada de decisão em questões relacionadas com esta temática. As mesmas poderiam ser utilizadas por vários tipos de utilizadores, porém, com especial relevância para entidades governamentais, partidos políticos e entidades privadas.

1.2. Objetivos de Investigação

A principal finalidade do presente Trabalho Final de Mestrado consiste em desenvolver uma solução de *Business Intelligence* (BI) de modo a analisar um conjunto de dados relativos ao movimento migratório da imigração em Portugal e, consequentemente, fornecer ajuda na tomada de decisão para assuntos relativos a este âmbito. Com recurso à ferramenta de visualização de dados, o intuito é criar painéis de visualização que disponibilizem informação relevante de uma forma clara, concisa e apelativa.

Neste sentido, pretende-se responder aos seguintes objetivos da investigação:

- Construir uma solução de *Business Intelligence* que permita suportar a tomada de decisão nas questões relacionadas com a imigração em Portugal:
 - Apresentar uma solução conceptual de BI;
 - Validar a exequibilidade do sistema.

Durante a realização do projeto, a solução escolhida será desenvolvida e serão apresentados *dashboards* recorrendo às técnicas de *data visualization* para uma melhor análise dos dados e, por conseguinte, uma melhor compreensão da complexa realidade da imigração. A solução será desenvolvida com recurso a uma ferramenta líder no mercado que permite transformar dados em informação útil e cuja visualização e análise pode ser feita de uma forma interativa.

1.3. Abordagem Metodológica

Para a realização do presente projeto foi desenvolvido um estudo tendo por base a abordagem metodológica *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Esta metodologia foi desenvolvida em 1996 por meio de um esforço conjunto das empresas DaimlerChrysler, SPSS e NCR, cuja finalidade era fornecer suporte na resposta a qualquer problema, independentemente da indústria (Chapman et al., 2000; Costa & Aparicio, 2020).

Sendo um dos processos de *Data Mining* (DM), CRISP-DM foi escolhido devido ao facto de ser de fácil compreensão e aplicação e tanto poder ser utilizada por especialistas em *data mining* como por quem não possui muita experiência nesta área. *Data Mining* (mineração de dados em português), representa um termo utilizado para descoberta de conhecimento a partir de grandes quantidades de dados (Turban, Sharda, Delen, 2010) identificando as estruturas e padrões neles presentes (Hand & Adams, 2015).

A metodologia CRISP-DM é apresentada em termos de um modelo com um processo hierárquico e cíclico composto por seis etapas, cuja sequência não é rígida. Estas etapas são: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* e *deployment* (Chapman et al., 2000), conforme se encontra representado na Figura 1. Dada a sua flexibilidade, CRISP-DM possibilita um processo iterativo onde é comum

recuar e avançar entre as etapas, consoante as necessidades de retificar ou ajustar a sequência do processo. Assim, as setas ilustradas na Figura 1 apenas exemplificam as direções e dependências mais comuns entre as fases.

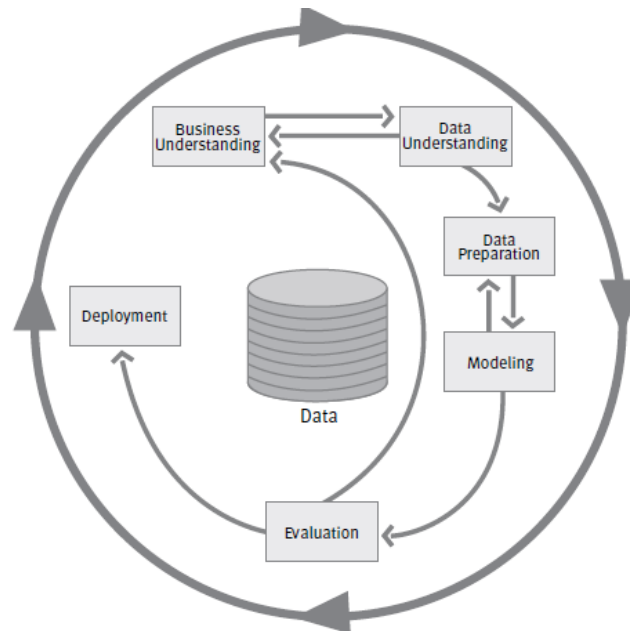


Figura 1 – Fases do modelo CRISP-DM (Fonte: Chapman et al., 2000)

Os mesmos autores descrevem cada uma das fases da seguinte forma:

Business Understanding - diz respeito à compreensão do negócio e à definição do objetivo do projeto. Para melhor compreender o problema é fundamental estudar o negócio onde este se encontra inserido. Depois, pode ser transformado num problema de DM e ser definido um plano para alcançar o objetivo do projeto.

Data Understanding – nesta fase são apresentados 4 passos para descobrir o problema e encontrar padrões nos dados. Estes passos são: recolha dos dados iniciais, descrição dos dados, exploração dos dados e a verificação da qualidade dos mesmos.

Data Preparation – consiste na preparação dos dados para serem tratados em ferramentas de *data mining*. Nesta etapa é selecionado o conjunto final de dados a serem usados, através de processos de limpeza, construção, integração e formatação.

Modeling – aborda a aplicação de técnicas e parametrizações aos dados de forma a obter um modelo que represente conhecimento.

Evaluation – refere-se à revisão e avaliação do processo já desenvolvido para a conceção do modelo. É também analisado de que forma o modelo contribui para o objetivo em questão.

Deployment – esta última fase é dedicada às atividades de organização dos resultados e na disponibilização dos mesmos para o utilizador final. Esta fase pode ser tão simples como gerar um relatório ou tão complexa como implementar um processo rentável de *data mining* em toda a empresa (Chapman et al., 2000).

1.4. Estrutura do Projeto

O presente Trabalho Final de Mestrado é composto por sete capítulos. No primeiro, é feito o enquadramento teórico do tema a ser estudado, são definidos os objetivos de investigação, assim como, é apresentada a abordagem metodológica que será aplicada durante a elaboração do projeto.

O segundo capítulo é relativo à revisão da literatura onde serão abordados dois temas distintos: a imigração em Portugal e o *Business Intelligence*, com destaque aos principais conceitos associados a este.

No capítulo três, é apresentada a aplicabilidade da abordagem metodológica ao tema em questão de forma a permitir a criação de uma solução de BI no contexto da imigração em Portugal. As fases da aplicação da metodologia são detalhadamente descritas de modo a alcançar os objetivos previamente definidos.

Os resultados obtidos são apresentados em forma de *dashboards* no quarto capítulo acompanhados de uma detalhada descrição de cada um deles.

No capítulo cinco, são expostas as conclusões, verificando o sucesso do projeto, nomeadamente o cumprimento dos seus objetivos. Também neste tópico são referidas algumas limitações sentidas ao longo da elaboração do projeto, assim como, as perspetivas futuras de investigação.

Por fim, nos últimos dois capítulos são incluídas as referências bibliográficas e anexos.

2. Revisão da Literatura

Nesta secção são abordados dois temas distintos: primeiramente é apresentada a realidade da Imigração em Portugal e posteriormente é descrito o *Business Intelligence*, assim como, alguns conceitos relacionados a este.

2.1. Imigração em Portugal

Segundo o Observatório das Migrações, no relatório “Imigração em Números de 2019”, 2019 foi o ano em que Portugal registou o maior número de imigrantes de sempre. No entanto, e apesar de apresentar um saldo migratório positivo, quando comparado com os restantes países europeus, Portugal é um dos que regista menor número de imigrantes demonstrando ainda ser pouco atrativo.

Conceptualmente, *imigrante* é definido como sendo alguém que se desloca entre fronteiras e/ou se fixa num país que não o seu de origem por um período superior a um ano (Oliveira & Gomes, 2019). Segundo o Glossário sobre Migrações¹, um imigrante permanente é alguém a quem foi legalmente atribuída a autorização de se fixar no país de acolhimento.

Em termos estatísticos, *imigrante* representa um conceito bastante inclusivo uma vez que se pode referir à sua naturalidade, nacionalidade ou até mesmo à naturalidade dos seus ascendentes (Huddleston, Niessen, & Tjaden, 2013). Os mesmos autores referem ainda que os *imigrantes* são definidos, nas estatísticas internacionais, consoante a disponibilidade de variáveis nas bases de dados nacionais que os permitam reconhecer como tal. Verifica-se, portanto, não haver um consenso de informação estatística internacional quanto à contabilização deste conjunto de indivíduos. Em Portugal, o principal critério utilizado para estimar a população imigrante e as suas características é o da nacionalidade, devido aos dados estatísticos e administrativos existentes serem pouco abrangentes. Apesar de ser um critério objetivo, torna-se também bastante limitativo uma vez que possibilita apenas uma aproximação ao universo de imigrantes no país. É, assim, importante realçar que, também a nível nacional, a informação estatística

¹ Elaborado pela Organização Internacional para as Migrações.

e administrativa existente não é suficiente para descrever com precisão a realidade da imigração em Portugal (Oliveira & Gomes, 2019) .

2.1.1. Tendências e Dinâmicas da Imigração em Portugal

Apesar dos padrões migratórios serem bastante diversos em todo o mundo, os autores do livro “*The Age of Migration*”- Castles & Miller (2009), identificaram um conjunto de tendências que podem ser consideradas gerais. Segundo Peixoto (2009) as migrações para Portugal a partir dos finais dos anos 90 apresentam algumas características bastante semelhantes quando comparados com outros fluxos migratórios recentes mas, evidentemente, com ritmos, caminhos e contextos diferentes, consoante os países. Algumas destas tendências são:

- Globalização: cada vez mais países são afetados pelo fenómeno da migração internacional, sendo que os países de imigração tendem a receber nacionais de países mais distintos, o que aumenta a diversidade no âmbito económico, social e cultural;
- Diferenciação de migrantes: diversos tipos de migração existentes, como a migração laboral, refugiados, reunião familiar, entre outros.;
- Aceleração: aumento dos movimentos migratórios;
- Feminização da migração laboral: aumento da migração feminina no contexto da migração laboral, sendo que no passado a maioria da migração feminina ocorria com o intuito de reunião familiar;
- Aumento da politização da migração: diversas políticas nacionais e internacionais são afetadas pela migração.

Peixoto (2009) refere ainda que foram sentidas alterações significativas nas migrações internacionais em Portugal entre 1980 e 2009. Estas alterações podem ser descritas destacando três fluxos principais: os migrantes vindos de África, sobretudo das ex-colónias Portuguesas, onde Portugal tem tido cada vez mais importância no papel de recetor; o aumento da presença dos imigrantes Brasileiros de classe social média-baixa; e, por fim, uma notória diversificação no que diz respeito aos países de origem dos imigrantes do Leste da Europa.

Nos anos 80 os imigrantes portugueses aumentaram consideravelmente. Na década seguinte o crescimento tornou-se ainda mais acentuado, apresentando uma taxa de crescimento anual de 7%, sendo que a percentagem de estrangeiros a residir em Portugal passou de 1.1% em 1990 para 1.9% em 1999 (Góis & Marques, 2018). De acordo com Fonseca (2008) este crescimento da imigração foi resultado de fatores internacionais como a queda do muro de Berlim, adesão à Comunidade Económica e Europeia, acordo de Schengen e também o progresso da economia portuguesa.

O início da década seguinte foi assinalado com a implementação de novas regras de autorização de permanência no território nacional o que, por sua vez, levou a um aumento da presença de residentes estrangeiros, sobretudo provenientes da Europa de Leste (Góis & Marques, 2018). Porém, em 2004 assistiu-se à estabilização desta evolução como consequência do fim das obras públicas, do declínio da atividade económica e também introdução de novas políticas migratórias Fonseca (2008). Os autores (Góis & Marques, 2018) referem que nos anos seguintes existiu uma alteração na estrutura da imigração em Portugal acompanhada de um crescimento significativo de entradas de cidadãos estrangeiros. Esta alteração coincide com entrada de um elevado número de imigrantes ucranianos que passam a integrar o grupo das três nacionalidades mais representativas em Portugal, juntamente com os imigrantes brasileiros e cabo-verdianos. Em conjunto, estes três grupos, passaram a constituir mais de metade do total de estrangeiros legalmente residentes no território nacional.

No período entre 2010 e 2015 verificou-se uma redução no número de residentes estrangeiros, resultado de uma crise económica e financeira que afetou o país. Como consequência desta crise, o decréscimo de oportunidades de trabalho levou à diminuição de novas entradas e ao aumento das saídas dos estrangeiros residentes. Por outro lado, esta diminuição também está relacionada com as aquisições de nacionalidade portuguesa por um elevado número de residentes estrangeiros (Oliveira & Gomes, 2019).

O acesso à nacionalidade constitui um importante fator tanto na integração da população estrangeira como na perspetiva do abrandamento do decréscimo e do envelhecimento populacional (Reis De Oliveira, Gomes, & Santos, 2017). As novas leis de aquisição de nacionalidade portuguesa sofreram alterações no ano 2006 onde inúmeros estudos destacaram Portugal como sendo um dos países da União Europeia (UE) mais

liberais e inclusivos na forma como concede a nacionalidade aos residentes estrangeiros (Huddleston, Bilgili, Joki, & Vankova, 2015). A aquisição de nacionalidade permite aos imigrantes uma melhor integração e envolvimento na sociedade de acolhimento, acesso a empregos com remunerações mais elevadas, bem como, acesso a uma melhor educação (Reis De Oliveira et al., 2017). Em Portugal, em termos percentuais os motivos para a aquisição da nacionalidade são: naturalização (70%), casamento ou união de facto há mais de três anos com nacional português (16%), atribuição originária (9%) e outros motivos (5%) (Ribeiro et al., 2019).

Os dados mais recentes refletem um notório aumento da população estrangeira residente, com destaque aos fluxos de migração laboral bem como de migrações de reformados e estudantes (Góis, 2019). Os números sugerem que em 2019 o aumento deste conjunto de indivíduos foi de aproximadamente 23% quando comparado com o ano anterior (Machado, Reis, Esteves, Sousa, & Rosa, 2020).

No que respeita à distribuição da população estrangeira no território nacional, existe um notório destaque para a escolha dos distritos do litoral. Mais de dois terços dos estrangeiros residentes em Portugal optaram por viver nos distritos de Lisboa, Setúbal e de Faro (Oliveira & Gomes, 2019). A escolha destas áreas advém sobretudo do seu desenvolvimento económico e uma oferta de emprego mais elevada quando comparada com os restantes distritos. No caso de Setúbal, a presença de elevado número de residentes estrangeiros é justificada pela influência de importantes zonas industriais. Relativamente ao distrito de Faro, a preferência deve-se ao facto de ser uma zona turística, em particular, durante a época balnear (D. Rodrigues, Correia, Pinto, Pinto, & Cruz, 2014)

2.1.2. Fatores *Push-Pull*

Existem vários fatores que podem levar a imigrar e diversas características para a escolha do país de destino. Uma das abordagens mais comuns relativa a este âmbito é a combinação de fatores “push-pull” que afetam diretamente a migração.

Segundo o autor Rosenberg (2020), “em termos geográficos, os fatores push-pull são aqueles que afastam as pessoas de um local e as atraem para outro”. Os fatores push são frequentemente fortes e maioritariamente negativos ou prejudiciais, isto porque

estimulam as pessoas a deixar os seus pontos de origem para se estabelecerem noutra lugar. Por outro lado, os fatores pull são positivos visto que incentivam e atraem as pessoas a imigrar para novos lugares, na maioria dos casos à procura de condições de vida melhores das que tinham até então.

Como exemplos de fatores push pode ser referido o desemprego, salários baixos, fome, pobreza, perseguições políticas, religiosas ou de certos grupos culturais, guerras civis e serviço militar obrigatório. Os fatores pull, por sua vez, estão relacionados com a procura de imigração no país de destino, como por exemplo a abundância de empregos, salários mais atrativos que o país de origem, segurança e liberdade pessoal, económica e religiosa, oportunidades de educação e melhor qualidade de vida em termos de saúde (Berg, 2004; Rosenberg, 2020).

Deste modo, pode ser concluído que Portugal é de certa forma mais atrativo para os imigrantes que aqui residem relativamente aos seus países de origem dado que apresenta melhores condições a nível económico, político e social.

2.1.3. Fatores económicos afetados pela imigração

As migrações constituem uma das mais relevantes influências sobre a transformação social em todo o mundo (Góis & Marques, 2018). Para além do impacto direto que exercem sobre o número e a estrutura da população de um determinado país, existem inúmeras outras consequências que advêm deste fenómeno. Estas tanto podem ser positivas como negativas, tanto podem ocorrer a curto como a longo prazo. Os fluxos migratórios afetam diretamente a economia de um país tendo um impacto notório nas suas taxas de crescimento, nas relações comerciais e balanças de pagamentos, na procura de educação, dos serviços, de saúde e infraestruturas sociais. Por outro lado, as condições políticas, sociais e culturais também são altamente afetadas (Cholezas & Tsakoglou, 2009). “Os fluxos migratórios (...) são capazes de acentuar, acelerar ou desacelerar processos de ordem económica, jurídica, política, social ou cultural” (Góis & Marques, 2018: p.126).

São variadas as contribuições favoráveis que advém da presença de população estrangeira em Portugal, nomeadamente, o impacto positivo no mercado laboral, derivado do potencial dos recursos humanos que os imigrantes representam, na educação, com a presença de elevado número de jovens, filhos de imigrantes, e também na segurança social devido às contribuições efetuadas por este conjunto de indivíduos (Oliveira & Gomes, 2018). No estudo desenvolvido por Góis (2019) é descrito de que forma os imigrantes contribuem para o desenvolvimento do país, apontando alguns pilares fundamentais.

No que diz respeito às contribuições políticas e culturais, o autor refere que a contribuição tem sido bastante positiva, apesar de não receber a devida atenção. Realça ainda, que a presença de pessoas com diferentes origens promove o progresso cultural que pode ser observado na literatura, música, dança e também na gastronomia.

Relativamente à demografia, observa-se que a população imigrante em Portugal “é tendencialmente mais jovem que a população portuguesa, concentrando-se em grupos etários mais jovens e em idades ativas e férteis” (Oliveira & Gomes, 2019: p. 56). Desta forma, a presença de imigrantes em Portugal desempenha um papel importante para atenuar os efeitos do decréscimo da população e do envelhecimento desta (Góis, 2019). Por outro lado, esta contribuição por parte dos imigrantes, sobretudo em idade ativa, resolve o problema do desequilíbrio presente no mercado de trabalho em vários setores (Góis, 2019). Segundo Góis (2019) e Oliveira & Gomes (2018) “19% dos imigrantes residentes regulares em Portugal estão incluídos no setor primário; 25% no secundário e 56% no setor terciário”. O aumento dos trabalhadores imigrantes compensou o decréscimo da mão-de-obra portuguesa sobretudo nos setores dos serviços e turismo (Góis, 2019).

Segundo o estudo desenvolvido pelo Faustino & Peixoto, (2009) o impacto da imigração no comércio português é claramente positivo. Os dados sugerem que as políticas de imigração que facilitam a admissão dos imigrantes favorecem a valorização das suas competências e promovem o empreendedorismo entre estes favorecendo, assim, a balança comercial e o padrão comercial.

Destaca-se ainda a importância do empreendedorismo como estratégia de integração laboral por parte da população estrangeira que proporciona também vantagens para a

economia portuguesa. As iniciativas de empreendedorismo podem ser vistas como uma forma de reduzir o desemprego através da criação de novos postos de trabalho que tanto podem ser ocupados por estrangeiros como pela população portuguesa. Adicionalmente, podem trazer alguma inovação para as empresas portuguesas como projetos inovadores, novas ideias e tecnologias atuais (Coutinho, Oliveira, Soares, & Sanchez, 2008).

2.2. *Business Intelligence*

2.2.1. Conceito de *Business Intelligence*

O ambiente empresarial sofre de constantes alterações e atualizações onde competir torna-se, cada vez mais, uma tarefa difícil. As organizações necessitam de estar cientes dessas alterações e serem bastante criativas na forma como operam. É fundamental que sejam ágeis e tomem as decisões estratégicas e operacionais de forma rápida e oportuna para responder às necessidades do mercado (Turban et al., 2010). Na atual era de informação o constante surgimento de uma elevada quantidade de dados impulsionou a necessidade de análises robustas com recurso a Sistemas de Informação (SI) de modo a entender e investigar os seus impactos (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016). A implementação de soluções de BI tem sido uma boa resposta a estes desafios na medida em que pode contribuir para uma melhor compreensão e análise de dados, transformando-os em informações uteis que sirvam de auxílio nas tomadas de decisão.

O conceito de *Business Intelligence* foi apresentado pela primeira vez em 1958 pelo investigador da IBM (Luhn, 1958) como sendo a capacidade de compreender as inter-relações dos factos através da definição de futuras ações com o intuito de alcançar um objetivo em concreto. No início dos anos 90 o analista do *Gartner Group* referiu o BI como sendo um conjunto de métodos e conceitos cujo propósito consistia em melhorar a tomada de decisão utilizando o conhecimento em sistemas (Cebotarean & Titu, 2010).

Segundo Turban et al. (2010) “*Business Intelligence* é um termo abrangente que combina arquiteturas, ferramentas, bases de dados, ferramentas analíticas, aplicações e metodologias.” Os mesmos autores defendem que o principal objetivo do BI reside em fornecer acesso interativo aos dados permitindo a manipulação dos mesmos, por vezes

em tempo real, auxiliando deste modo os gestores de negócios e analistas efetuar análises apropriadas.

Por outro lado, Zeng, Xu, Shi, Wang, & Wu (2006) descrevem *Business Intelligence* como sendo “um processo de recolha, tratamento e difusão de informação que tem como objetivo a redução de incerteza na tomada de decisões estratégicas”.

Em síntese, o processo de *BI* consiste em criar soluções que transformam dados em informação e conhecimento que, posteriormente, dão origem a decisões estratégicas. Estas, por sua vez, são convertidas em ações concretas (Olszak & Ziemba, 2007). O processo descrito pode ser observado na Figura 2.

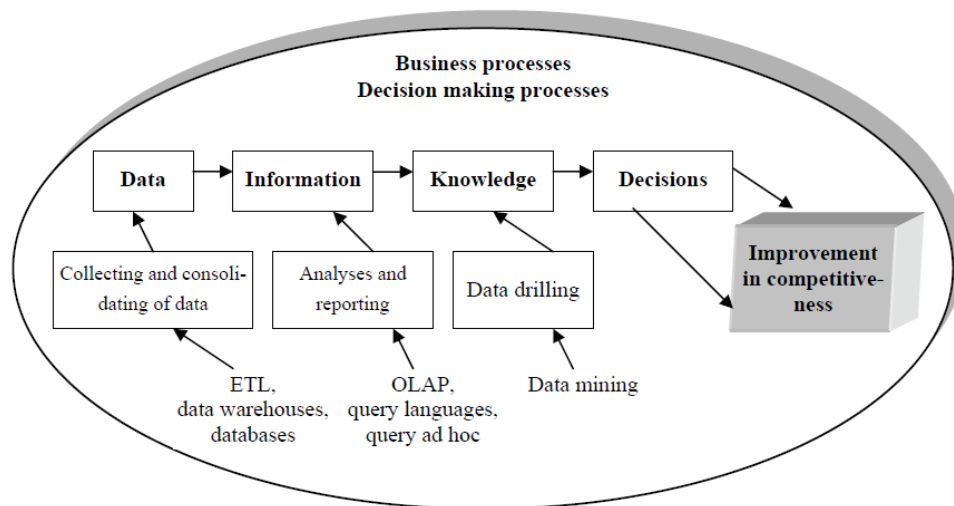


Figura 2 – O Papel dos sistemas de BI na tomada de decisão. (Fonte: Olszak & Ziemba, 2007)

Para uma melhor compreensão do BI é fundamental entender alguns conceitos associados ao mesmo. *Data Warehouse* (DW) é um dos mais referenciados pelos autores pois está diretamente relacionado às bases de dados (BD), presentes em qualquer organização, mas com o comum problema de estarem dispersas por diversas áreas. Este conceito representa um repositório de dados históricos e correntes que contêm relevância para a organização (Turban et al., 2010, Caetano & Costa, 2014). É muitas vezes referido como um “armazém de dados”, isto é, um repositório onde os dados possuem variadas origens, organizados e agrupados num esquema e guardados num local único de forma a fornecer apoio à gestão (Han, Kamber, & Pei, 2013). *Data Warehouses* são construídas através de processos de limpeza, integração, transformação, carregamento dos dados e

também uma atualização periódica destes (Han et al., 2013). Para a criação da estrutura de uma DW é frequentemente utilizada a modelação multidimensional que proporciona uma estrutura de base de dados de fácil compreensão e utilização. A modelação multidimensional será abordada e descrita no tópico 2.2.2.

No que respeita à análise das *Data Warehouses* existem várias ferramentas que ajudam na monitorização, perceção e integração da informação. Uma das mais conhecidas é a utilização das técnicas de OLAP (*Online Analytical Processing*). Estas técnicas permitem construir respostas rápidas para análises multidimensionais dos dados, isto é, analisar a informação de diferentes perspetivas. OLAP refere-se, portanto, à execução de análises complexas das informações guardadas em *Data Warehouses* (Pasatcha, 2007). Segundo os autores do livro “*Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*”, Moss & Atre (2002), as tecnologias OLAP criam novas informações com base em transformações robustas e cálculos efetuados utilizando os dados já existentes.

Uma vez que as fontes de dados, por vezes, são variadas e os dados nem sempre se encontram estruturados, é necessário aplicar um conjunto de transformações de forma a construir estruturas compatíveis com os requisitos das ferramentas de BI (Kimball & Caserta, 2004). A este processo é dado o nome de ETL (*Extract, Transform and Load*). Segundo os autores (Boateng, Singh, & Greeshma, 2012) ETL pode ser definido como o processo de armazenamento que consiste na extração, transformação e carregamento de dados, integrados e limpos, em *Data Warehouses*. A extração é o primeiro passo deste processo que consiste na leitura de dados das várias fontes. De seguida, é feita a transformação dos dados extraídos através da sumarização da informação, correção dos possíveis erros e exclusão de duplicados e campos irrelevantes. Por fim, é realizado o carregamento dos dados já transformados em *Data Warehouses* (Kimball & Ross, 2013). ETL é uma das componentes mais importantes de BI pois a qualidade dos dados conseguida neste processo está fortemente ligada à qualidade de dados das outras componentes (Dayal, Castellanos, Simitsis, & Wilkinson, 2009).

2.2.2. Modelação Multidimensional

A modelação multidimensional é uma técnica de estruturação de informação bastante intuitiva para o utilizador (Kimball, Ross, Thornthwaite, Mundy, & Becker, 2008). É utilizada para construir estruturas de dados de fácil compreensão e com capacidade para a otimização do desempenho de processamento de *queries* no sistema (Moody & Kortink, 2003). Este modelo é composto por uma tabela de factos central e um conjunto de tabelas de dimensões (Kimball et al., 2008) que se relacionam através de um modelo de dados e são apresentadas em forma de esquema (Kazi, Kazi, & Radulovic, 2012). Segundo Ferrari & Russo (2016), um modelo de dados consiste num conjunto de tabelas interligadas entre si através de relações entre as chaves identificadoras, isto é, colunas partilhadas entre cada conjunto de tabelas. A chave primária identifica cada linha na tabela e a chave estrangeira referencia um valor que é encontrado noutra tabela (Microsoft, 2020a). Segundo Allen & Terry (2005) as relações entre as tabelas possuem uma determinada direção e cardinalidade. As relações operam através da correspondência de dados em colunas-chave (normalmente colunas ou campos) que partilham o mesmo nome em ambas as tabelas (Microsoft, 2020a). No que diz respeito às direções é necessário saber quais são as tabelas que serão ligadas: tabela de origem (são aquelas onde é obtido o valor a ser pesquisado na outra tabela) e tabela de destino (corresponde à tabela onde é efetuada a pesquisa). Os mesmos autores realçam ainda que a cardinalidade das relações se refere à quantidade de vezes que cada valor da coluna chave aparece nas colunas relacionadas. Os possíveis tipos de cardinalidade entre as tabelas são: um para um, um para muitos e muitos para muitos, sendo que o tipo das relações depende de como as colunas relacionadas são definidas (Allen & Terry, 2005; Microsoft, 2020a)

A tabela de factos é o centro do esquema e está relacionada ao assunto ou âmbito do negócio que pretende ser analisado. É composta por um conjunto de atributos (factos) e também um conjunto de chaves estrangeiras que relacionam a tabela com diferentes tabelas de dimensões (Santos & Ramos, 2009).

As tabelas de dimensões fornecem a base para analisar os elementos da tabela de factos. Tipicamente são compostas por atributos que respondem a questões “quem?”, “o quê?”, “quando?”, “onde?”, “como?” e “porquê?” relativamente aos componentes da tabela de factos (Moody & Kortink, 2003).

Os autores do livro “*Data Mining: Concepts and Techniques*”, Han et al., (2013), referem que o modelo de dados mais utilizado para DW é o modelo multidimensional. Este pode ser construído em forma de estrela (*star schema*), floco de neve (*snowflake schema*) ou constelação de factos (*fact constellation*):

Star Schema – é uma estrutura de representação de dados única e altamente regular. É composta por uma tabela de factos central e várias tabelas de dimensões cujos atributos estão ligados à tabela de factos. Num esquema em estrela, as tabelas de dimensões apenas podem estar ligadas à tabela de factos e nunca entre elas (Ferrari & Russo, 2016). Ao aplicar esta estrutura aos dados é possível evitar ambiguidade nas relações do modelo, minimizando a quantidade de repetições do mesmo valor numa coluna ou em várias linhas, tornando o modelo mais eficiente (Ari & Tolvanen, 2019). A sua esquematização pode ser observada na Figura 3.

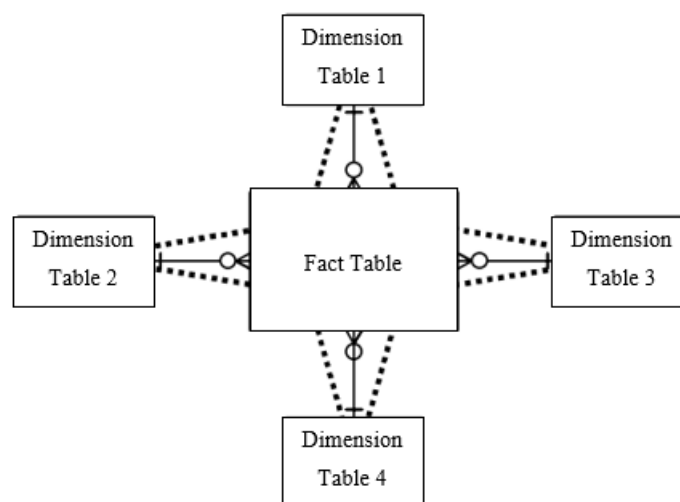


Figura 3 – Estrutura genérica de um esquema em Estrela. (Fonte: Moody & Kortink, 2000)

Snowflake Schema – a estrutura deste modelo tem por base o esquema em estrela, onde algumas tabelas de dimensão são normalizadas, dividindo ainda mais os dados em tabelas adicionais, como pode ser verificado na Figura 4. Contudo, a normalização pode dificultar a interpretação por parte do utilizador final e, também, uma redução no desempenho no processamento de consultas (*queries*) (Han et al., 2013).

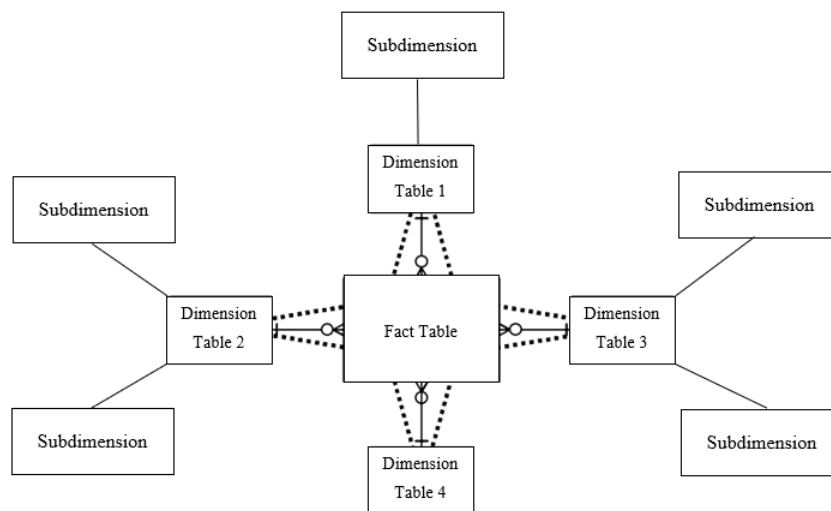


Figura 4 – Estrutura genérica de um esquema em Floco de Neve. (Fonte: elaboração própria com base em Moody & Kortink, 2000)

Fact Constellation – ou *Galaxy Schema*, é outra possível estrutura de modelação de dados. Esta é composta por várias tabelas de factos que partilham uma ou várias dimensões (Santos & Ramos, 2009). Traduz-se na junção de vários esquemas em estrela pelo que, é frequentemente referida como esquema em galáxia (Han et al., 2013). A sua representação pode ser verificada na Figura 5.

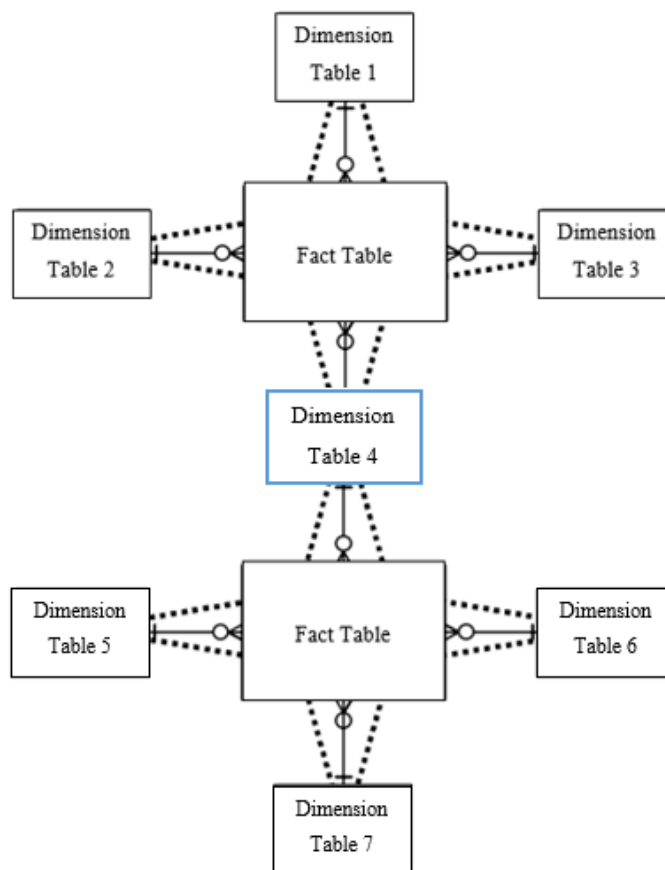


Figura 5 – Estrutura genérica de um esquema em Constelação. (Fonte: elaboração própria com base em Moody & Kortink, 2000 e Santos & Ramos, 2009)

A escolha entre os modelos apresentados depende sobretudo das propriedades do sistema que pretende ser desenvolvido. É possível aplicar o esquema em floco de neve parcialmente, apenas para algumas dimensões, ou até mesmo dentro de uma dimensão (Jensen, Pedersen, & Thomsen, 2010).

2.2.3. Data Visualization

O aumento crescente do volume de dados disponíveis originou, nas empresas, uma necessidade de melhorar os seus processos de tomada de decisão, baseando-os em fontes ricas de informação, provocando, desta forma, um uso cada vez maior de *dashboards* (Malik, 2005). Estes ajudam a ter uma visão instantânea dos aspetos mais críticos do negócio e, por outro lado, ajudam a identificar e melhorar as possíveis falhas. Um *dashboard* (painel de visualização) é uma apresentação visual de um conjunto de

informações agregadas e organizadas numa tela única, relevantes para atingir um ou vários objetivos (Few & Edge, 2007). Eckerson (2011) realça que os *dashboards* permitem ao utilizador monitorizar a informação e atuar sobre os acontecimentos assim que estes acontecem, aumentando desta forma, a eficiência e melhorando os resultados. A visualização da informação num *dashboard* deve ajudar o utilizador a compreender melhor dados complexos e as causas que poderão estar na raiz de um determinado problema. Assim sendo, o resultado pretendido com um *dashboard* não é apenas fornecer informação útil, mas também criar um processo de reflexão no seu utilizador (Skorka, 2017).

Firican (2017) define um conjunto de características para a criação de um *dashboard* eficaz. O autor salienta que o *dashboard* deve ser atual e fiável, ou seja, os dados utilizados na sua elaboração devem estar sempre atualizados, embora a frequência da sua atualização varie consoante as necessidades do negócio e do propósito deste. É importante incluir, também, detalhes acerca da última atualização dos dados, assim como, as respetivas fontes. Os títulos e as descrições, por sua vez, devem ser claros e sucintos. Em caso de existirem vários *dashboards* com um *layout* semelhante, os títulos acompanhados de uma breve descrição servirão de apoio na sua distinção. Desta forma, o utilizador final terá maior facilidade em encontrar a informação que é relevante para si. Por outro lado, é recomendada a utilização de meios visuais em vez de tabelas ou *pivot tables*. Estes meios visuais podem ser: um gráfico de barras (útil quando o objetivo é mostrar uma ou várias séries de dados tais como, o tempo, o produto, a receita ou a região); um gráfico de linhas (vantajoso para acompanhar alterações em vários conjuntos de dados); mapas (quando se pretende apresentar dados geográficos ou espaciais); a aplicação de filtros (proporciona ao utilizador um processo mais interativo); legendas (sempre que for utilizada mais que uma cor ou formato o uso das legendas passa a ser fundamental). Por último, é mencionada a acessibilidade aos *dashboards* na medida em que, se possível, estes devem estar disponíveis para consulta, idealmente na web, para que qualquer utilizador autorizado os possa facilmente visualizar.

Para Costa & Aparicio (2019), o sucesso de um *dashboard* depende da informação nele representada. Esta deve estar extremamente bem organizada, resumida e customizada consoante as necessidades e objetivos do utilizador. Deve, também, ser apresentada através de meios concisos onde a comunicação é feita de forma clara e direta.

Diferentes tipos de decisões necessitam de diferentes formas de apresentação da informação (Costa & Aparicio, 2019). Os mesmos autores propõem um modelo em árvore de decisão que permite, com base nos dados disponíveis, identificar quais são os tipos de gráficos que devem ser utilizados. Por outro lado, Abela (2006) sugere outra forma que ajuda a selecionar a visualização mais adequada consoante os dados existentes, pois defende que uma má escolha do gráfico pode levar a resultados menos precisos. O guia desenvolvido pelo autor encontra-se representado em anexos (Figura 19).

2.2.4. Software

Para a elaboração do projeto foi utilizado *Microsoft Power BI* que é uma aplicação da *Microsoft* que permite, de uma forma bastante simples e intuitiva, transformar dados em informação útil através das ferramentas de visualização. *Power BI* é uma solução *SSBI* (*Self Service BI*), isto é, permite construir e ativar uma solução de forma rápida (Microsoft, 2020b). A ferramenta foi destacada pelo 13º ano consecutivo como líder no relatório “2020 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms” conforme representado na Figura 20 (em Anexo).

Power BI permite acesso rápido e fácil às bases de dados com várias origens, localizados na *cloud* ou localmente. Possibilita a visualização dos dados numa tela única, garantindo que estes estejam sempre atualizados. As ferramentas de análise desta aplicação são bastante intuitivas, permitindo que o processo de tomada de decisão seja mais instruído.

3. Aplicação da Metodologia CRISP-DM

A metodologia CRISP-DM foi escolhida para a elaboração do presente projeto devido à sua flexibilidade, eficiência e ampla aplicabilidade, podendo ser ajustada a qualquer setor da indústria onde o utilizador pode definir a tecnologia que pretende utilizar (Wirth & Hipp, 2000). Foram seguidas as seis etapas da metodologia adotada, efetuando-se a análise a cada uma delas. Numa primeira fase são definidos os objetivos do projeto. De seguida, é feita a extração dos dados necessários para a elaboração do projeto. Os dados são trabalhados e transformados numa terceira fase, onde também é selecionado o conjunto final de dados a ser utilizado. Após a limpeza e transformação, num quarto estágio do processo, é feita a modelação, isto é, aplicação de modelos

escolhidos sobre os dados selecionados. É também nesta fase que são construídos os painéis de visualização. Por fim, é feita a avaliação do resultado obtido fazendo a comparação dos objetivos definidos na primeira fase.

3.1. *Business Understanding*

O fenómeno da imigração é uma temática bastante importante pois pode trazer, para Portugal, mais valias a nível socioeconómico, cultural e político. Como referido na revisão da literatura, os fluxos migratórios influenciam de forma direta as taxas de crescimento, as relações comerciais, a procura de educação, serviços e saúde (Cholezas & Tsakoglou, 2009).

Contudo, a informação disponível sobre este tema é complexa, exige muita pesquisa e análises detalhadas. Maioritariamente é apresentada em relatórios extensos, cuja interpretação requer tempo e esforço. Por outro lado, os dados sobre a imigração encontram-se dispersos por várias fontes, o que torna a pesquisa ainda mais demorada. Assim sendo, pretende-se com este projeto fornecer auxílio nas análises da realidade imigratória em Portugal, através da disponibilização de *dashboards* visuais e dinâmicos que sumarizam e destacam apenas a informação relevante. Com base nestes painéis de visualização o utilizador poderá fazer a sua avaliação e interpretação de uma forma mais eficaz e tomar as decisões de uma forma mais rápida.

Apesar de não ter um utilizador final definido, os *dashboards* poderão ser consultados por diversas entidades. Seguem-se alguns exemplos dos possíveis utilizadores:

- Entidades Governamentais – a consulta dos *dashboards* por entidades governamentais pode facilitar a tomada de decisão no que diz respeito à elaboração das políticas associadas à imigração ou à caracterização social de Portugal;
- Partidos Políticos – através da análise dos painéis de visualização poderá ser possível perceber quais são os países de origem cujos indivíduos adquirem mais a nacionalidade portuguesa e utilizar esta informação para desenvolver estratégias de modo a atrair mais votos através de elaboração de propostas que sejam do interesse desse conjunto de indivíduos;

- Entidades Autárquicas – por meio da observação dos painéis de visualização, destacar as nacionalidades mais presentes de forma a desenvolver medidas de integração para esse grupo de pessoas (por exemplo, fornecendo aulas de português);
- Organizações não governamentais (ONG) – com a utilização dos *dashboards* será possível perceber junto de que grupo de indivíduos existe uma maior necessidade de atuação;
- Utilizadores académicos – tais como estudantes, investigadores ou professores, podem utilizar os *dashboards* de forma a obter uma melhor visualização da informação sobre a imigração em Portugal e, por conseguinte, aumentar a qualidade do estudo ou da investigação;
- Entidades Privadas – nomeadamente empreendedores que pretendem desenvolver o seu negócio direcionado para uma determinada nacionalidade. Como exemplo, pode ser referida uma loja com produtos típicos de um determinado país, pois frequentemente, estes indivíduos têm preferência por produtos do seu país natal que não são comercializados em Portugal. Com base na informação dos *dashboards* poderá ser possível não apenas a identificação do público alvo, como também, a construção da mensagem de comunicação a transmitir.

3.2. Data Understanding

A segunda fase do processo refere-se à recolha e análise dos dados. Nesta fase também é verificada a sua qualidade e são definidas e descritas as variáveis.

Os dados foram extraídos em formato Livro do Excel de cinco fontes de dados abertos:

- Instituto Nacional de Estatística (INE) – autoridade de Portugal que produz e publica informação estatística oficial;
- PORDATA - base de dados que contem estatísticas certificadas sobre Portugal;

- Portal de estatística do Serviço de Estrangeiros e Fronteiras (SEFSTAT) - portal que disponibiliza informação fiável e de qualidade relativamente a população estrangeira em Portugal;
- ANumbeo – base de dados colaborativa sobre cidades e países do Mundo;
- Indexmundi – base de dados que contem estatísticas detalhadas dos países, assim como, gráficos e mapas compilados de várias fontes.

Após recolha, os dados foram agregados num único ficheiro Excel de modo a que a sua exportação para a aplicação *Power BI* seja efetuada de forma mais simples.

Os dados abertos, segundo a definição da (Agência para a Modernização Administrativa, 2016), são dados públicos, normalmente em bruto, de acesso fácil e universal, sem restrições de utilização e em formatos que possam ser facilmente trabalhados por aplicações informáticas. No Guia da Introdução aos Dados Abertos (Agência para a Modernização Administrativa, 2016) estão refletidos os benefícios vindos da utilização de dados abertos. Para além dos ganhos no que diz respeito à transparência, há igualmente um forte potencial económico associado. Estes benefícios estão refletidos tanto no setor privado como na Administração Pública. No entanto, e talvez devido à sua rápida disseminação, muitos dos dados abertos disponíveis já se encontram bastante trabalhados, isto é, não estão em bruto como nos diz a definição da AMA. Estão normalmente organizados em quadros de forma a que a sua leitura seja mais fácil pelo que, por vezes, têm de ser desmontados e reorganizados dependendo do tratamento pretendido.

O primeiro conjunto de dados é referente à população estrangeira com estatuto legal de residente. Segundo Instituto Nacional de Estatística (2019) a população é definida como sendo “conjunto de pessoas de nacionalidade não portuguesa com autorização ou cartão de residência, em conformidade com a legislação de estrangeiros em vigor. Não inclui os estrangeiros com situação regular ao abrigo da concessão de autorizações de permanência, de vistos de curta duração, de estudo, de trabalho ou de estada temporária, bem como os estrangeiros com situação irregular”. Para este conjunto de indivíduos foi recolhido o número total de pessoas por ano, no período entre 2008 e 2019, assim como, o número de Mulheres (M) e de Homens (H) integrados neste conjunto. Também foi

extraída a informação acerca da distribuição deste grupo pelo território nacional, mais precisamente por distrito, também fazendo a distinção entre M, H e Total.

Outro conjunto de dados é relativo à origem da população estrangeira residente em Portugal, ou seja, foi extraído o número de pessoas de cada nacionalidade também no período entre 2008 e 2019.

Adicionalmente, foi reunida a informação acerca da população estrangeira que adquiriu a nacionalidade portuguesa. Desta forma, foi possível saber o número de atribuições de nacionalidade portuguesa por cada país relativamente ao período entre 2008 e 2018, visto que a informação do ano de 2019 ainda não se encontra disponível.

No que respeita aos países de origem da população estrangeira residente em Portugal, foi também recolhida informação sobre o Indicador de Qualidade de Vida, o Indicador de Custo de Vida e o Produto Interno Bruto (PIB) *per capita* do país em questão. Conforme (Numbeo, 2020) o Indicador de Qualidade de Vida “é uma estimativa da qualidade de vida geral usando uma fórmula empírica que tem em conta o indicador de poder de compra (maior é melhor), indicador de poluição (menor é melhor), relação preço de casa para rendimento (menor é melhor), indicador de custo de vida (menor é melhor), indicador de segurança (maior é melhor), indicador de cuidados de saúde (maior é melhor), indicador de tempo de comuta no trânsito (menor é melhor) e indicador de clima (maior é melhor). O Indicador de Custo de Vida (excluindo renda) é definido segundo (Numbeo, 2020) como “um indicador relativo de preços de bens de consumo, incluindo mercearia, restaurantes, transportes e serviços. Indicador de Custo de Vida não inclui despesas de alojamento como rendas ou créditos.” É importante referir que este indicador é relativo a Nova Iorque o que significa que “para Nova Iorque, cada indicador deve ser 100(%)”. Se um país tiver, por exemplo, um indicador de restaurantes de 130%, reflete que, em média, as refeições nos restaurantes nesse país serão 30% mais caras que em Nova Iorque” (Numbeo, 2020). Por último, o PIB *per capita* foi retirado de (IndexMundi, 2020) e “mostra o PIB em uma base de paridade do poder de compra dividida pela população”. É um indicador económico que relaciona o crescimento de uma economia com a sua população.

A análise do conjunto de dados recolhidos permitiu identificar as variáveis e o seu tipo. Esta informação encontra-se resumida na Tabela 1, juntamente com a descrição das variáveis.

Nome da variável	Descrição	Tipo de variável
Ano	Período em Análise	Quantitativa Discreta
Género	M = Mulher H = Homem	Qualitativa Nominal
Número de Pessoas (pop. estrangeira)	Refere-se à população estrangeira residente em Portugal	Quantitativa Discreta
Distrito	Distritos de Portugal	Qualitativa Nominal
Número de pessoas (pop. estrangeira por distrito)	Refere-se à população estrangeira em Portugal por distrito	Quantitativa Discreta
País de Origem	País de origem da população estrangeira residente	Qualitativa Nominal
Número de pessoas (Origem dos Imigrantes)	Refere-se à população estrangeira residente em Portugal por país de origem	Quantitativa Discreta
Indicador de Qualidade de Vida	Percentagem estimada da qualidade de vida geral num país	Quantitativa Contínua
Indicador de Custo de Vida	Percentagem estimada do custo de vida num país relativo a Nova Iorque	Quantitativa Contínua
PIB <i>per capita</i>	Produto Interno Bruto dividido pela quantidade de habitantes de um país.	Quantitativa Contínua

Tabela 1 – Descrição das variáveis.

A qualidade dos dados foi verificada através da comparação direta entre várias fontes. Neste sentido, uma vez que os dados foram extraídos de várias fontes, e apesar destas serem fidedignas, nem sempre aparentam ser consistentes. Tal facto deve-se à inconsistência nacional na forma como a população estrangeira é contabilizada, como referido na revisão de literatura. Não obstante, é de realçar que as divergências verificadas não terão impacto na construção dos painéis de visualização nem na análise da informação.

Esta fase corresponde ao *Extract* do processo ETL (*Extract, Transform, Load*) referido no estado da arte.

3.3. Data Preparation

Na terceira etapa do processo, procedeu-se ao tratamento e limpeza dos dados extraídos. Para facilitar o uso da ferramenta *Power BI* e uma correta leitura dos dados, estes foram compilados num ficheiro único de Excel que servirá de base de dados para o projeto. Inicialmente, a base de dados criada foi explorada a fim de identificar variáveis que não acrescentavam valor bem como variáveis com valores nulos e por consequência, estas foram eliminadas.

O conjunto final de dados encontra-se organizado em quadros. As fontes de dados abertos disponibilizam as informações neste formato de modo a facilitar a sua leitura por parte do utilizador. Porém, para trabalhar com os dados em *Power BI* é necessário proceder a sua reorganização e aplicação de um conjunto de transformações. Estes dois passos intermédios serão feitos dentro da ferramenta. Na figura que se segue pode ser observado de que forma foi feita a importação dos dados (1), a escolha das *tabs* do ficheiro Excel (2) e a transformação dos dados (3).

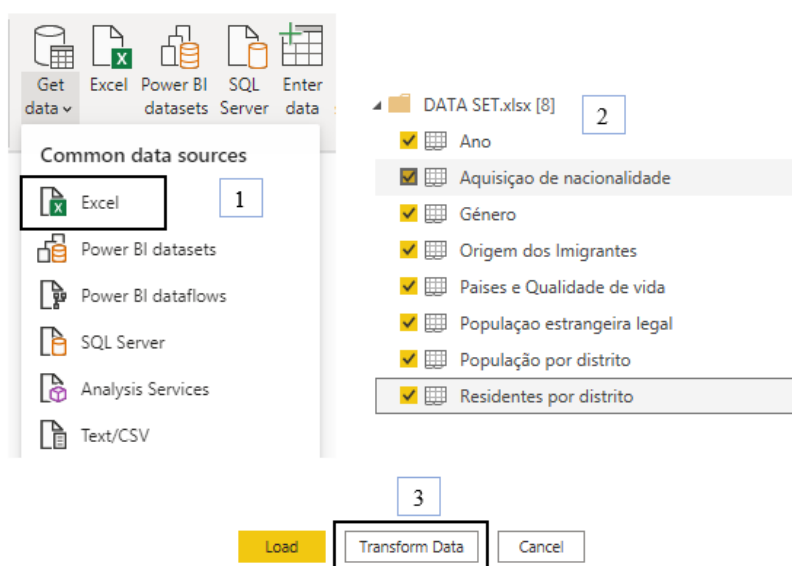


Figura 6 – Importação, Seleção e Transformação dos dados em *Power BI*.

Ao selecionar a opção “*Transform Data*” a aplicação abre uma nova janela - *Power Query Editor* - onde serão feitas as restantes transformações.

Para uma apropriada leitura dos dados em *Power BI* é necessário que estes estejam organizados em tabelas, isto é, cada linha tem de corresponder a um registo independente

de uma ocorrência. Para tal, os dados serão transformados em ambiente *Power Query Editor*, com a aplicação da função “*unpivot table*” como pode ser observado na Figura 7.

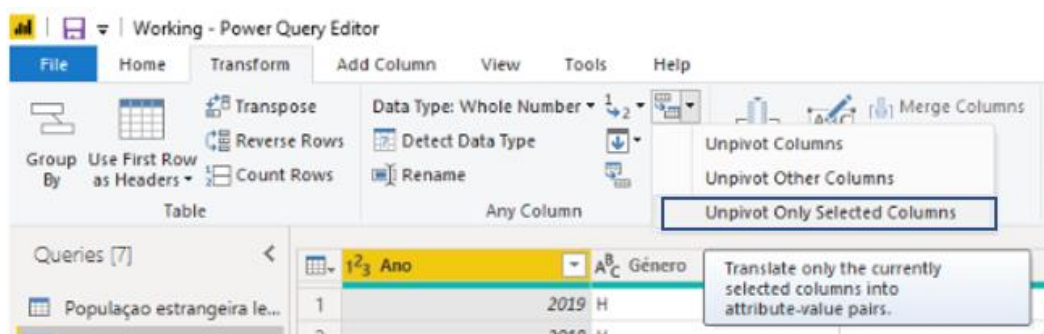


Figura 7 – Transformação dos dados para uma estrutura em tabela.

Para além de transformar a estrutura dos dados em tabela, é importante garantir que cada coluna esteja na formatação correta, de acordo com o tipo de dados nela presentes. Por exemplo, a coluna correspondente ao “Ano” deve ser do tipo “*whole number*”, a coluna relativa ao “Género” deve ser do tipo “*text*” e assim sucessivamente.

Alguns dos conjuntos de dados necessitam de eliminação de linhas e colunas em branco, assim como, linhas que contenham títulos ou descrições. Este passo pode ser observado na Figura 8.

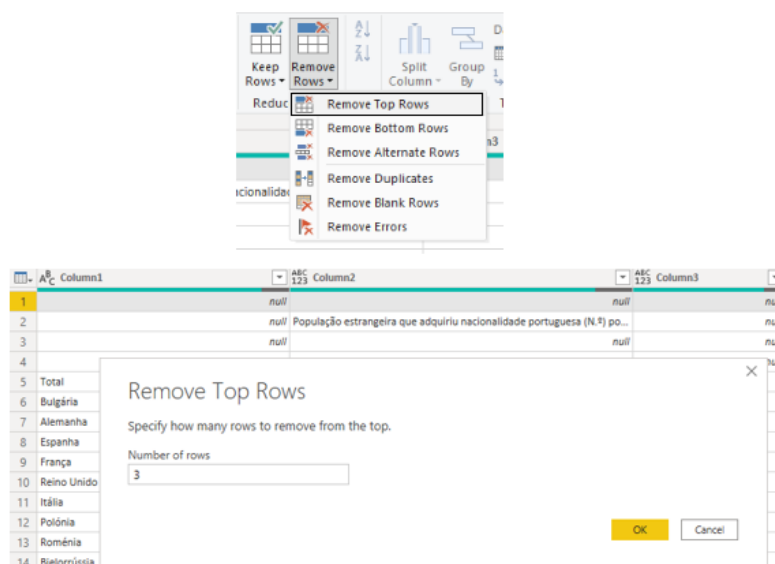


Figura 8 – Eliminação de linhas.

Adicionalmente, para a criação de uma das visualizações cuja representação será feita em *Flow Map* foi necessário acrescentar a coluna “Destino” à tabela “Origem dos

Imigrantes”. Este tipo de mapas “mostram o movimento de objetos de uma localização para outra, como por exemplo o número de pessoas numa migração” (Phan, Xiao, Yeh, Henrahan, & Winograd, 2005). A criação da coluna adicional é demonstrada passo a passo na Figura 9.

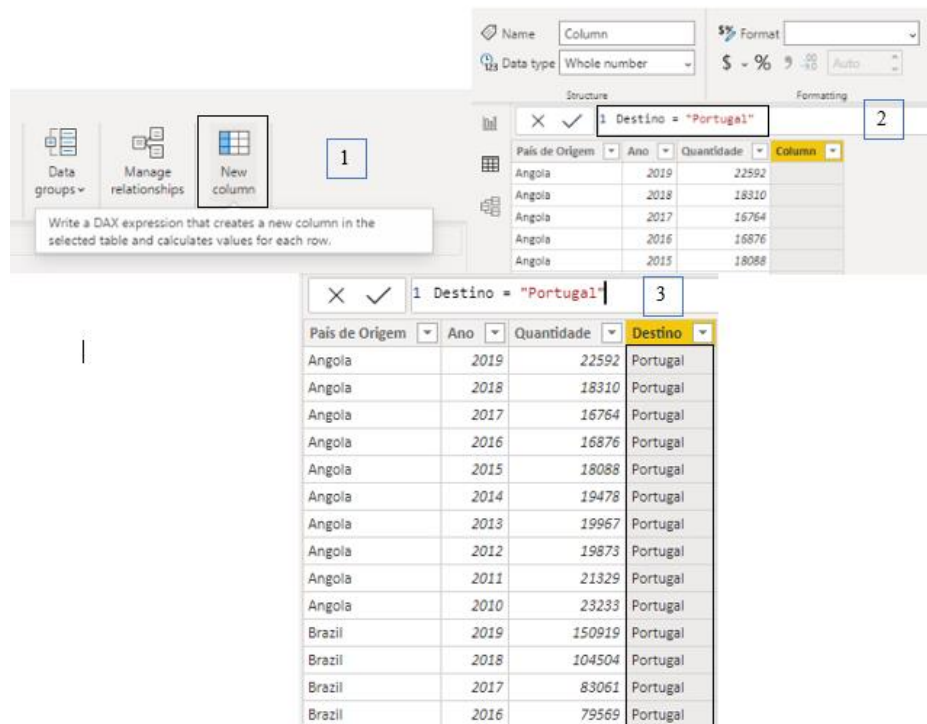


Figura 9 – Criação de colunas adicionais.

Após a aplicação deste conjunto de transformações a várias tabelas de dados será, finalmente, efetuado o “Load” de modo a dar continuidade ao processo.

É de realçar que sempre que a base de dados criada no Excel for atualizada, os dados em *Power BI*, assim como, todos os elementos visuais ligados a estes, também serão automaticamente atualizados ao utilizar a função “Refresh”.

O processo de limpeza descrito e o carregamento dos dados equivalem ao “Transform” e “Load” do processo ETL.

3.4. Modeling and Visualization

Após a seleção do conjunto final de dados e do seu *upload* na aplicação, é necessário cruzar os dados presentes em diferentes tabelas, ou seja, identificar os atributos que tenham em conjunto e, deste modo, criar as relações entre as várias tabelas. Uma

vantagem eficaz da aplicação utilizada é conseguir criar estas relações automaticamente. Desta forma, o utilizador apenas necessita de verificar se as relações criadas são viáveis e, caso exista alguma tabela que não tenha nenhuma ligação, criá-la manualmente.

Durante a execução desta etapa, a tabela “Género” não se encontrava relacionada com nenhuma outra, pelo que as relações associadas a esta foram criadas manualmente ao seleccionar a opção *Manage relationships* → *New* (Figura 10 (a)). De seguida, basta averiguar quais são as tabelas que contêm os atributos da tabela “Género” e seleccioná-las uma a uma, como está representado na Figura 10 (b). A cardinalidade é detetada automaticamente pelo *Power BI*.

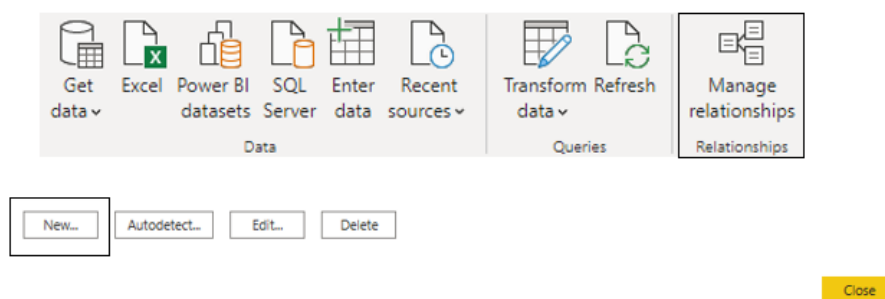


Figura 10 (a) – Criação manual das relações em *Power BI*.

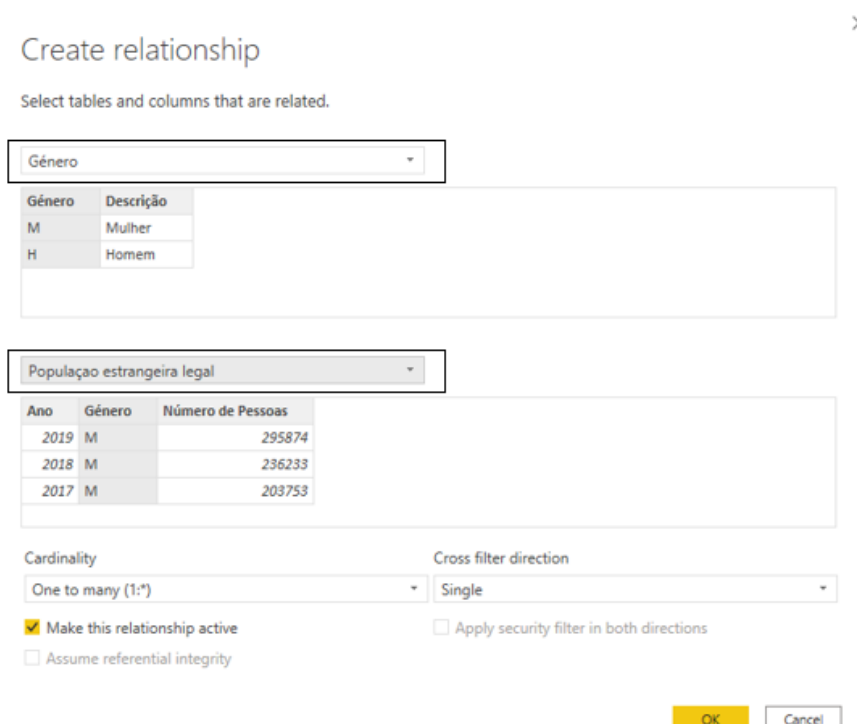


Figura 10 (b) - Criação manual das relações em *Power BI*.

Num modelo de dados existem dois tipos de tabelas, as tabelas de factos e as tabelas de dimensões, como referido na revisão da literatura. Na Figura 11, pode ser visualizado o modelo relacional presente nos dados, constatando a existência de 4 tabelas de factos: Residentes por distrito (1), População estrangeira legal (2), Origem dos Imigrantes (3) e Aquisição de nacionalidade (4), sendo que as restantes correspondem às tabelas de dimensões, pois categorizam e qualificam as informações das tabelas de factos. Uma vez que existem várias tabelas de factos que partilham algumas das dimensões, estamos perante um Esquema em Constelação.

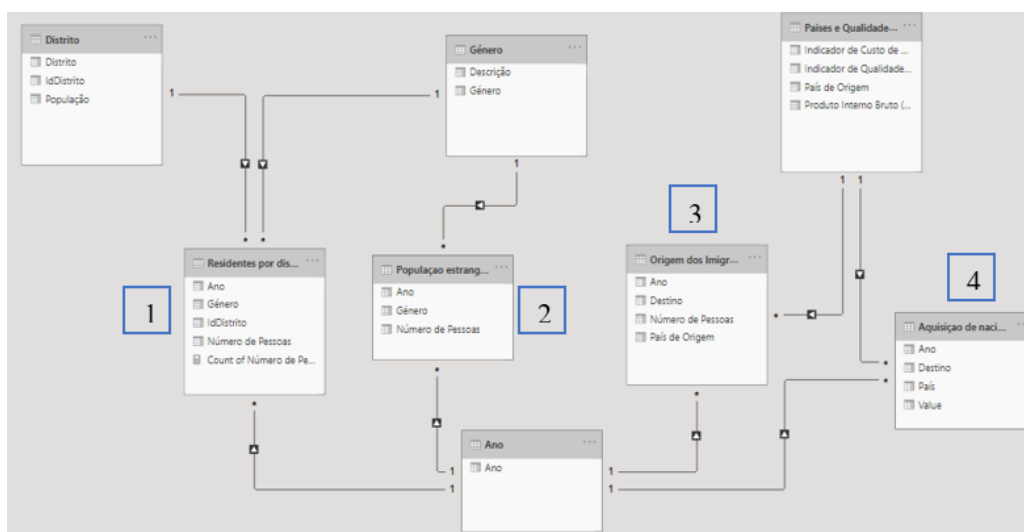


Figura 11 – Modelo Relacional.

A Tabela que se segue expressa o resultado das tabelas criadas, bem como, das relações existentes entre estas, complementando a representação gráfica da Figura 13.

Tabela de Facto	Tabela de Dimensão	Atributo em comum
Residentes por distrito	Distrito	IdDistrito
	Género	Género
	Ano	Ano
População estrangeira legal	Género	Género
	Ano	Ano
Origem dos Imigrantes	Países e Qualidade de vida	País de Origem
	Ano	Ano
Aquisição de nacionalidade	Países e Qualidade de vida	País de Origem
	Ano	Ano

Tabela 2 – Modelo Relacional.

Após a criação das relações entre as tabelas e da verificação da sua viabilidade é possível avançar para a próxima etapa que consiste no desenho dos elementos de visualização.

Como foi referido no estado da arte, um *dashboard* é uma tela única que agrega um conjunto de informações e apresenta-as de uma forma visualmente apelativa. É importante ter em conta que os painéis de visualização devem apresentar apenas os elementos mais relevantes de um determinado tema, consoante o objetivo do utilizador.

Tendo em conta os dados recolhidos para a elaboração deste projeto, os *dashboards* foram desenvolvidos de forma a reunir os tópicos relevantes para a tomada de decisão em questões relacionadas com a imigração em Portugal, utilizando as visualizações disponíveis no *Power BI*. Deste modo, foram desenvolvidos cinco *dashboards* que relatam diferentes cenários. A seleção dos elementos de visualização que compõem cada um dos *dashboards* foi feita de acordo com o guia sugerido por Abela (2006), referido na revisão de literatura.

Na Figura 12 encontra-se representada a esquematização da solução desenvolvida, isto é, os *dashboards* elaborados. Primeiramente, foi criada uma janela introdutória que é composta por cinco botões, um para cada painel, dando ao utilizador a possibilidade de seleccionar o que deseja analisar de seguida. O primeiro *dashboard* aborda a evolução da população estrangeira residente em Portugal no período entre 2008 e 2019, recorrendo à tabela “População estrangeira legal”. É composto por um gráfico de linhas, um sectograma e dois filtros – um que afetará o gráfico de linhas, e o segundo que será relativo ao sectograma. No segundo painel, é exposta a distribuição deste conjunto de indivíduos por distrito, no mesmo período de referência. Assim sendo, é também utilizado um gráfico de linhas, acompanhado de um filtro, desta vez utilizando como suporte as tabelas “Residentes por distrito” e “Distrito”. Adicionalmente, foi criado um sectograma, recorrendo à tabela “Género” que será afetado por um outro filtro. A terceira página mostra as nacionalidades mais representativas em Portugal, através do uso de um mapa de fluxo, uma tabela e um filtro que, por sua vez, influencia todas as visualizações da página. Os três elementos de visualização têm por base a tabela “Origem dos Imigrantes”. O quarto *dashboard* é composto por quatro visualizações do tipo *Card* cujo conteúdo muda consoante a escolha feita no filtro. Este *dashboard* faz a comparação entre os três

tipos de indicadores económicos de Portugal e dos dez países mais destacados em Portugal no ano de 2019. A sua produção foi feita com base na tabela “Países e Qualidade de vida”. Por fim, o último painel de visualização foi construído recorrendo à visualização *Map*, *Table* e *Slicer* de forma a mostrar quais são os países que mais adquirem a nacionalidade portuguesa. Para a sua construção foi utilizada a tabela “Aquisição de nacionalidade”.

A descrição mais detalhada de cada um dos *dashboards* será feita no Capítulo 4, juntamente com as possíveis informações que o utilizador pode extrair desses *dashboards*.

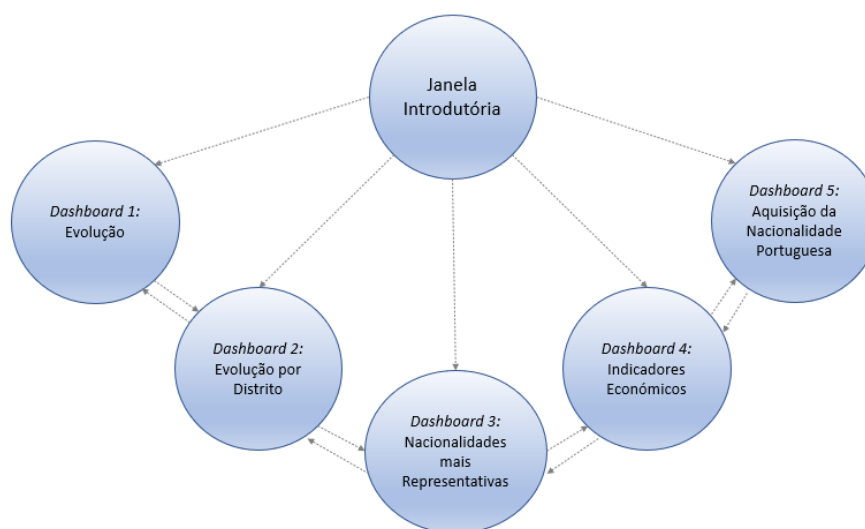


Figura 12 – Esquematização da Solução Desenvolvida.

3.5. Evaluation

A análise SWOT é um método eficaz para analisar situações correntes, tomar decisões e investigar possíveis escolhas futuras para selecionar a melhor estratégia. Tem em conta não apenas a perspetiva do momento, mas também oportunidades futuras (Dehghan, Mehrabi, & Fotouhi, 2013). Através desta análise é possível identificar as forças, fraquezas, oportunidades e ameaças a que o projeto está sujeito, o que ajuda a gerir o seu planeamento e também a sua implementação (Sabbaghi & Vaidyanathan, 2004). Por estas razões, este método foi o escolhido para avaliar o processo já desenvolvido.

De forma a obter uma avaliação preliminar do projeto, foi pedida a opinião a um profissional da área de *Power BI* cujas funções frequentemente passam por desenvolver

dashboards sobre a realidade da empresa onde trabalha. Sendo que possui bastante experiência na área, foi possível obter uma avaliação adequada e ponderada. Depois de lhe ter sido apresentada a solução desenvolvida, o profissional, indicou um conjunto de vantagens, desvantagens e também possíveis áreas de melhoria. Estes aspetos foram transformados numa análise SWOT, cujo resultado pode ser consultado na Tabela 3.

<p style="text-align: center;"><u>Forças</u></p> <ul style="list-style-type: none"> ○ Fácil compreensão da informação apresentada; ○ Possibilidade de agrupar várias informações num só painel de visualização; ○ <i>User friendly</i> e interativo; ○ Visualmente apelativo. 	<p style="text-align: center;"><u>Oportunidades</u></p> <ul style="list-style-type: none"> ○ Aplicabilidade a outras áreas de investigação, nomeadamente emigração e migrações internas; ○ Aumento da diversidade de visualizações dos <i>dashboards</i> através dos recursos visuais; ○ Divulgação de informação sobre a imigração em Portugal a utilizadores interessados, com recurso a <i>sites</i> e <i>blogs</i>.
<p style="text-align: center;"><u>Fraquezas</u></p> <ul style="list-style-type: none"> ○ Pouca diversidade dos dados; ○ Possíveis incongruências das fontes de dados; ○ Custo da implementação da tecnologia utilizada. 	<p style="text-align: center;"><u>Ameaças</u></p> <ul style="list-style-type: none"> ○ Surgimento de ferramentas concorrentes com maior capacidade de armazenamento e processamento de dados; ○ Surgimento de aplicações mais elucidativas para os utilizadores com menos experiência.

Tabela 3: Análise SWOT da Solução de BI.

3.6. *Deployment*

O projeto foi desenvolvido utilizando a ferramenta *Power BI Desktop* pelo que a disponibilização dos *dashboards* apenas é possível acedendo ao *desktop* onde estes foram guardados ou publicando na área pessoal da aplicação do usuário, contudo a sua consulta *online* não pode ser feita por nenhum outro indivíduo a não ser o proprietário da conta onde foi publicado. Não sendo possível a sua partilha devido às restrições da própria ferramenta, o presente Trabalho Final de Mestrado não contempla a implementação da última etapa da metodologia adotada. Porém, a finalidade do projeto reside no desenvolvimento de uma solução de BI que auxilie na tomada de decisão em questões associadas à imigração em Portugal, pelo que, o seu sucesso ou insucesso não dependerá deste passo.

4. Resultados

Neste capítulo pretende-se descrever e analisar os *dashboards* desenvolvidos previamente. Cada um destes *dashboards* relata um cenário diferente e contempla um conjunto diversificado de informação que pode ser facilmente detetada pelo utilizador.

Primeiramente, foi construída uma janela introdutória (Figura 13) constituída por um título e cinco subtítulos, cada um acompanhado de um botão incorporado nas imagens, permitindo ao utilizador escolher qual o *dashboard* que deseja explorar.



Figura 13 – Janela Introdutória.

De seguida apresentam-se os *dashboards* referidos na Janela Introdutória:

- “Evolução da População Estrangeira em Portugal (2008-2019)”

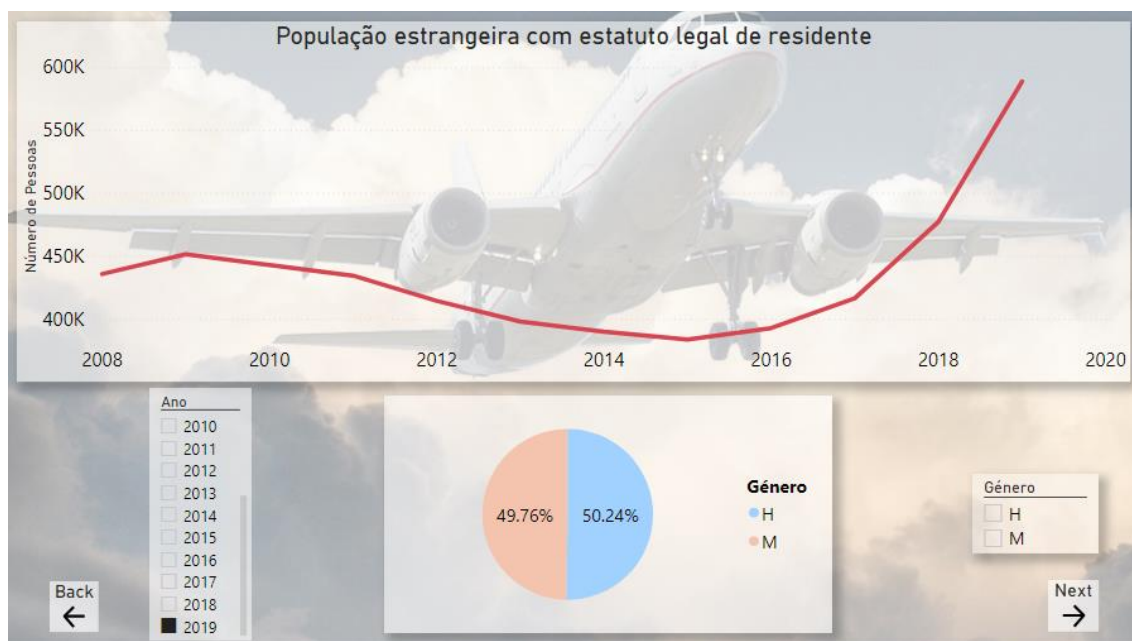


Figura 14 - Evolução da População Estrangeira em Portugal (2008-2019)

O painel tem como objetivo apresentar a evolução da população estrangeira com estatuto legal de residente em Portugal entre 2008 e 2019. Para saber o número exato de pessoas, em cada ano, basta arrastar o cursor para a área do gráfico de linhas e a informação será exibida. Para além da evolução geral, é possível aplicar o filtro “Género” que altera o gráfico mostrando a evolução do género, homem ou mulher, selecionado. O utilizador também tem disponível a informação relativa à percentagem da população por género para cada um dos anos ao utilizar o filtro “Ano”. Na Figura 21 (em Anexo) pode ser observado o exemplo da aplicação dos filtros. Assim, num único *dashboard*, foi possível mostrar elementos diferentes, o que poderá ser vantajoso para o utilizador visto que numa só tela tem disponível vários tipos de informação, não necessitando de realizar pesquisas morosas. Para avançar para o próximo *dashboard*, basta clicar no botão “Next”.

- “Evolução da População Estrangeira por Distrito (2008- 2019)”

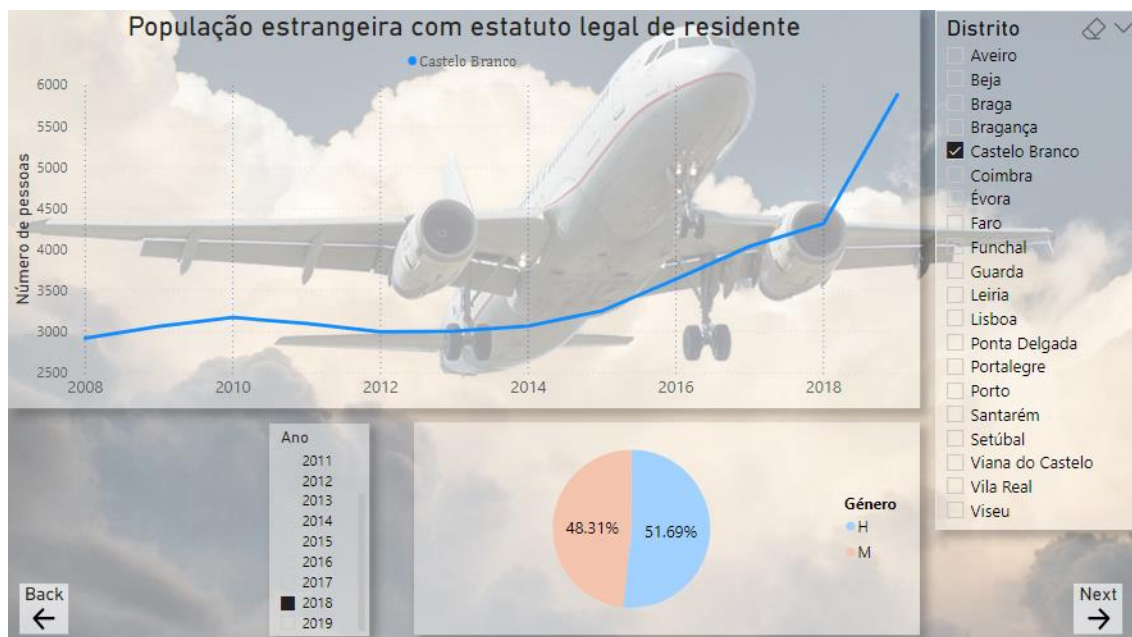


Figura 15 - Evolução da População Estrangeira por Distrito (2008- 2019).

Na página seguinte, é possível visualizar a distribuição da população estrangeira com estatuto legal de residente por distrito, assim como, a sua evolução no período de referência. Uma vez mais, movendo o cursor ao longo do gráfico o utilizador observa os valores para cada distrito consoante o ano que pretende consultar (Figura 22 em Anexo). Ao aplicar o filtro “Distrito”, é possível obter a evolução do número deste conjunto de indivíduos para o distrito selecionado, como se encontra representado na Figura 15. Caso seja do interesse do utilizador comparar a evolução entre dois ou mais distritos, a solução apresentada também o permite, ao escolher as opções desejadas no filtro. A sua representação pode ser encontrada na Figura 23 em Anexos. Adicionalmente, o utilizador pode comparar a evolução dos distritos com a evolução geral em Portugal mudando para a página anterior através do botão “Back”. Por fim, filtrando o ano escolhido pelo utilizador, encontra-se disponível para consulta a percentagem de cada género.

- “Nacionalidades mais representativas em Portugal”



Figura 16 – Nacionalidades mais representativas em Portugal.

O terceiro *dashboard* desenvolvido mostra quais são as nacionalidades que estão mais presentes em Portugal. O utilizador tem a possibilidade de saber, de forma praticamente imediata, quais foram as dez nacionalidades mais representativas em cada ano, ao aplicar o filtro “Ano” e visualizar os respetivos países no mapa. É de salientar que o mapa não é estático, permitindo ao utilizador aproximar e mover sempre que pretender.

Para além do mapa, é possível conhecer o número exato de pessoas vindas de cada um dos países de origem. Esta informação encontra-se apresentada por ordem decrescente e pode ser consultada na tabela situada no canto inferior direito da tela. Adicionalmente, caso o utilizador esteja interessado em saber quais foram os países menos representativos de um determinado ano deverá clicar na seta por baixo do título da segunda coluna da tabela, e a ordem inverte, evidenciando os dez países com menos população estrangeira em Portugal.

- “Comparação entre os Indicadores Económicos”



Figura 17 – Comparação dos Indicadores Económicos.

No quarto painel de visualização, o utilizador tem a possibilidade de comparar indicadores económicos referentes à qualidade de vida, custo de vida e PIB *per capita* entre Portugal e os dez principais países de origem dos imigrantes. Tal como pode ser verificado na Figura 17, do lado esquerdo são apresentados os indicadores referentes a Portugal, sendo que estes são valores de referência fixa. Do lado direito, achou-se pertinente acrescentar um filtro que permitisse a seleção de diferentes países mostrando, desta forma, os respetivos indicadores. Com base neste *dashboard* o utilizador pode efetuar várias análises no sentido de compreender a possível razão que levou estes imigrantes a escolher Portugal como destino.

- “Aquisição de Nacionalidade Portuguesa”



Figura 18 – Aquisição da Nacionalidade Portuguesa.

Por fim, o último *dashboard* foi elaborado com intuito evidenciar os países que mais adquirem a nacionalidade portuguesa. Deste modo, o utilizador tem a opção de escolher no filtro o ano que pretende analisar e de forma imediata as visualizações são atualizadas. O mapa destaca os principais dez países cujos nacionais adquiriram a nacionalidade portuguesa no ano selecionado, sendo que as *bubbles* nele posicionadas são proporcionais ao número de aquisições de nacionalidade por país (ou seja, maior tamanho corresponde a um maior número de pessoas que adquiriram a nacionalidade portuguesa). Adicionalmente, o utilizador pode saber qual foi o número exato de pessoas que adquiriram a nacionalidade por país de origem ao observar a tabela situada no canto superior esquerdo. Esta foi construída de forma a apresentar os países por ordem decrescente, contudo, caso o intuito seja saber os países que menos solicitam a nacionalidade portuguesa, basta clicar na seta situada por baixo do título da segunda coluna da tabela e a ordem é invertida, permitindo também a visualização dos países no mapa. À semelhança do *dashboard* “Nacionalidades mais representativas em Portugal”, o mapa aqui apresentado também é dinâmico, sendo possível a sua aproximação, afastamento e movimentação.

5. Conclusões, Limitações e Trabalho Futuro

5.1. Conclusões

O presente projeto teve como propósito desenvolver uma solução de *Business Intelligence* que suportasse a tomada de decisão em questões relacionadas com a imigração em Portugal. Para atingir este objetivo principal, foi necessário apresentar uma solução conceptual de BI, bem como, validar a exequibilidade do sistema.

Inicialmente, o processo da revisão da literatura permitiu examinar os conceitos mais relevantes para o desenvolvimento do projeto. Por um lado, foi importante entender algumas questões relacionadas com a imigração em Portugal. Por outro, os conceitos subjacentes ao *Business Intelligence* permitiram compreender de que forma seria desenvolvida a solução.

Para a execução do projeto foi seguida a adaptação da metodologia *Cross-Industry Standard Process for Data Mining* pondo em prática as suas etapas, com o intuito de atingir o objetivo final. É importante realçar que a metodologia escolhida foi adaptada ao conceito de *Business Intelligence* de forma a incorporar as técnicas *Data Visualization* no desenho da solução. Durante o desenvolvimento do presente projeto a solução conceptual foi criada aplicando os conhecimentos adquiridos na revisão da literatura. Relativamente à exequibilidade do sistema, este foi posto em prática através da aplicação dos dados recolhidos. O resultado é apresentado em forma de painéis de visualização que dispõem de conteúdo estruturado sobre a realidade imigratória em Portugal. Para a sua criação foi utilizada a ferramenta *Power BI Desktop* que permitiu transformar e estruturar os dados recolhidos, criar um modelo entre estes e construir os *dashboards* desejados. Sendo uma aplicação bastante intuitiva e *user friendly*, o processo da sua aprendizagem foi simples. Os *dashboards* desenhados disponibilizam a informação num formato visualmente apelativo cuja interpretação é fácil, permitindo ao utilizador uma tomada de decisão mais elucidativa.

Depois de garantir que os objetivos do projeto foram cumpridos com sucesso, isto é, a solução desenvolvida é funcional, os *dashboards* criados foram avaliados por um

profissional da área de *Power BI* que destacou um conjunto de aspetos positivos, assim como, algumas sugestões de melhoria. Estas informações foram analisadas e transformadas numa análise SWOT evidenciando forças, fraquezas, oportunidades e ameaças do sistema.

5.2. Limitações e Trabalho Futuro

Durante a criação do projeto foram sentidas e apontadas algumas limitações que, por vezes, dificultaram o seu desenvolvimento. Inicialmente, a fase da recolha de dados relativos à imigração em Portugal demorou mais que o esperado, pois os dados existentes eram escassos, levando a uma pesquisa mais aprofundada e, consequentemente, mais demorada. Por outro lado, já numa fase mais avançada, os dados recolhidos apresentaram ser pouco diversificados, dificultando o seu cruzamento em várias tabelas.

Por outro lado, a aplicação *Power BI Desktop* limitou o projeto na medida em que não foi possível disponibilizar os *dashboards* criados junto a um utilizador. Caso a publicação dos *dashboards online* permitisse a sua partilha, seria possível elaborar uma avaliação mais promenorizada.

Por consequência das limitações apresentadas, o projeto possui alguns aspetos que podem ser melhorados através de investigações futuras. Seguem-se, portanto, as principais sugestões neste âmbito:

- 1) Seria vantajoso proceder a recolha de mais dados relacionados com a imigração em Portugal. Caso a tarefa não seja possível recorrendo aos dados abertos, seria benéfico pedir às instituições como SEF ou INE, dados mais diversificados, garantindo o cruzamento com os dados já extraídos. Desta forma, os *dashboards* tornar-se-iam ainda mais completos e, consequentemente, as análises e interpretações seriam mais abrangentes.
- 2) Seria também importante disponibilizar os painéis de visualização em formato *online* através da integração destes em *sites* ou *blogs*, de modo a que qualquer utilizador tenha um acesso fácil.

6. Referências Bibliográficas

- Abela, A. V. (2006). Choosing a good chart - The Extreme Presentation(tm) Method. Retrieved September 17, 2020, from https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html
- AMA. (2016). *Guia de Introdução aos Dados Abertos*.
- Allen, S., & Terry, E. (2005). *Beginning Relational Data Modeling* (2nd ed.). Apress.
- Ari, H., & Tolvanen, A. (2019). *Developing an efficient business intelligence solution for day-to-day supply chain management: Case Pulp and Paper Industry Workshop*, tese de Mestrado, Lappeenranta-Lahti University of Technology
- Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues Of Big Data And Analytics In Networked Business. *Mis Quarterly*, 40(4), 807–818.
- Berg, H. Van den. (2004). *International Economics* (1st ed.). McGraw-Hill.
- Boateng, O., Singh, J., & Greeshma, P. S. (2012). Data Warehousing. *Business Intelligence Journal*, (July). Vol. 5 Num. 2, pp. 224-2354
- Caetano, T. V., & Costa, C. J. (2014). Data Warehousing num contexto de Sistemas Integrados. In *Atas da Conferência da Associação Portuguesa de Sistemas de Informação*, Vol. 12, pp. 186-199.
- Castles, S., & Miller, M. J. (1998). *The Age of Migration - International Population Movements in the Modern World* (2nd ed.). London: Macmillan.
- Cebotarean, E., & Titu, M. (2010). Business Intelligence. *Journal of Knowledge, Management, Economics and Information Technology*, 1(2), 101-113.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.
- Cholezas, I., & Tsakloglou, P. (2009). The economic impact of immigration in Greece: Taking stock of the existing evidence. *Journal of Southeast European and Black Sea Studies*. 9(1-2), 77-104.
- Costa, C., & Aparicio, M. (2019). “Supporting the decision on dashboard design charts.” *Proceedings of 254th The IIER International Conference*, 10–15.
- Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A Methodology to Boost Data Science. *Iberian Conference on Information Systems and Technologies, CISTI*, (June). doi: 10.23919/CISTI49556.2020.9140932.
- Coutinho, A. L., Oliveira, B., Soares, V., & Sanchez, S. M. (2008). *Immigrant entrepreneurship in Portugal : Factors that influence this professional choice and ASI ' s activity on foot of the results of the PEI Project*.
- Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). *Data Integration Flows for Business Intelligence*. In Proc. of the 12th International Conference on Extending Database Technology: Advances in Database Technology, pp. 1-11.
- Dehghan, A., Mehrabi, A., & Fotouhi, N. (2013). The necessity of establishing Business Intelligence competency centers for achievement of BI projects. *IKT 2013 - 2013 5th Conference on Information and Knowledge Technology*, 242–247.
- Eckerson, W. W. (2011). *Performance Dashboards: Measuring, Monitoring and Managing Your Business* (2nd ed.). Wiley.
- Faustino, H., & Peixoto, J. (2009). Immigration-Trade Links: The Impact of Recent Immigration on Portuguese Trade. *Working Papers Department of Economics*.
- Ferrari, A., & Russo, M. (2016). *Introducing Microsoft Power BI*. Microsoft Press.

- Few, S., & Edge, P. (2007). Dashboard Confusion Revisited. *Visual Business Intelligence Newsletter*. Perceptual Edge
- Firican, G. (2017). Best Practices for Powerful Dashboards. *Business Intelligence Journal*. vol. 2, num 2, 33-39.
- Fonseca, M. L. (2008). Imigração, Diversidade e Novas Paisagens Étnicas e Culturais. *Portugal. Percursos de Interculturalidade. Contextos e Dinâmicas*.
- Góis, P. (2019). *Migrações e desenvolvimento em Portugal. Avançar nas práticas: rumo à inclusão e coesão social*.
- Góis, P., & Marques, J. C. (2018). Retrato de um Portugal migrante: a evolução da emigração, da imigração e do seu estudo nos últimos 40 anos. *E-Cadernos CES*.
- Han, J., Kamber, M., & Pei, J. (2013). Data Mining: Concepts and Techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013* (3rd ed.).
- Hand, D. J., & Adams, N. M. (2015). Data Mining. *StatsRef: Statistics Reference Online*.
- Huddleston, Niessen, J., & Tjaden, J. T. (2013). *Using EU indicators of immigrant integration - EU Law and Publications*.
- Huddleston, T., Bilgili, Ö., Joki, A.-L., & Vankova, Z. (2015). Migrant Integration Policy Index. *MIPEX, CIDOB, MPG: Barcelona e Bruxelas*.
- IndexMundi. (2020). Comparação entre países - Produto Interno Bruto (PIB) per capita. Retrieved September 3, 2020, from <https://www.indexmundi.com/g/r.aspx?c=mr&v=67&l=pt>
- INE. (2019). População estrangeira com estatuto legal de residente. Retr. August 12, 2020, https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE&xlang=pt
- Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2010). Multidimensional Databases and Data Warehousing. In *Synthesis Lectures on Data Management* (Vol. 2).
- Kazi, L., Kazi, Z., & Radulovic, B. (2012). Data Warehouse Based Evaluation of Students' Achievements in Information Systems Education. *Proc. 35th Int. Conv. on Inf. Com. Tech, Electronics and Microelectronics (MIPRO 2012)*, 1377–1382.
- Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit*. Wiley.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley.
- Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., & Becker, B. (2008). *The Data Warehouse Life-cycle Toolkit* (2nd ed.). Wiley.
- Luhn, H. P. (1958). A Business Intelligence System. *IBM Journal of Research and Development*. 2(4), 314-319.
- Machado, R., Reis, S., Esteves, S., Sousa, P., & Rosa, A. P. (2020). Relatório de Imigração, Fronteiras e Asilo 2019. SEF.
- Malik, S. (2005). Enterprise Dashboards: Design and Best Practices for IT. In *John Wiley and Sons Inc* (Vol. 1). New Jersey: Wiley.
- Microsoft. (2020a). Como definir relações entre tabelas numa base de dados do Access. Retrieved September 9, 2020, from <https://docs.microsoft.com/pt-pt/office/troubleshoot/access/define-table-relationships>
- Microsoft. (2020b). Microsoft|Power BI. Retrieved September 8, 2020, from <https://powerbi.microsoft.com/pt-pt/>
- Moody, D., & Kortink, M. A. . (2000). From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design. *Proceedings of the Int. Workshop on Design and Management of Data Warehouses (DMDW'2000)*.
- Moody, D. L., & Kortink, M. A. R. (2003). From ER Models to Dimensional Models:

- Bridging the Gap between OLTP and OLAP Design, Part I. *Business Intelligence Journal*. Summer, 7-24
- Moss, L. T., & Atre, S. (2002). *Business Intelligence Roadmap : The Complete Project Lifecycle for Decision-Support Applications*. Boston: Addison-Wesley.
- Numbeo. (2020). Custo de Vida. Retrieved September 3, 2020, from <https://pt.numbeo.com/custo-de-vida/página-inicial>
- Oliveira, C. & Gomes, N. (2018). *Indicadores de integração de imigrantes: relatório estatístico*. Observatório das migrações, Lisboa
- Oliveira, C. R., & Gomes, N. (2019). *Indicadores de Integração de Imigrantes Relatório Estatístico Anual 2019*. Observatório das migrações, Lisboa
- Oliveira, C. R. de, & Gomes, N. (2018). *Migrações e saúde em números: o caso português*. Observatório das Migrações, ACM, IP.
- Olszak, C. M., & Ziemba, E. (2007). Approach to building and implementing Business Intelligence systems. *Interdisciplinary Journal of Information, Knowledge, and Management*. 2(1), 135-148.
- Pasatcha, P. (2007). *Techniques and Process Based Decision Support in Business Intelligence System*. International Conference on e-Business 2007 (INCEB 2007) pp. 22 - 23
- Peixoto, J. (2009). New migrations in Portugal: Labour markets, smuggling and gender segmentation. *International Migration*. 47(3), 185-210
- Phan, D., Xiao, L., Yeh, R., Henrahan, P., & Winograd, T. (2005). Flow Map Layout. In IEEE Symposium on Information Visualization, 2005. INFOVIS 2005. pp. 219-224. IEEE.
- Reis De Oliveira, C., Gomes, N., & Santos, T. (2017). *Acesso à Nacionalidade Portuguesa: 10 anos da Lei em Números*. Retrieved from https://www.om.acm.gov.pt/documents/58428/440932/Caderno+Estatístico+1+Nacionalidade+FINAL_b.pdf/787fbb3d-d8ea-4daa-922d-5cbefec49ff4
- Ribeiro, M. J., Estrela, J., Rosa, A. P., Cruz, M., Miranda, S., Sousa, P., & Machado, R. (2019). *Sef 2018 Relatorio Imigração Fronteiras Asilo*. SEF
- Richardson, J., Sallam, R., Schlegel, K., Konz, A., & Sun, J. (2020). 2020 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms. Gartner, Inc.: Stamford, CT, USA, 1-58.
- Rodrigues, D., Correia, T., Pinto, I., Pinto, R., & Cruz, C. (2014). Um Portugal de Imigrantes: exercício de reflexão sobre a diversidade cultural e as políticas de integração *Investigação às práticas*, 4(1), 82-105.
- Rodrigues, T. F., & Ferreira, S. D. S. (2014). Portugal e a globalização das migrações . Desafios de segurança. *População e Sociedade*, 22(CEPESE, Porto), 137–155.
- Rosenberg, M. (2020). Push-Pull Factors in Immigration. Retrieved April 28, 2020, from <https://www.thoughtco.com/push-pull-factors-1434837>
- Sabbaghi, A., & Vaidyanathan, G. (2004). SWOT Analysis and Theory of Constraint in Information Technology Projects. *Information Systems Education Journal*.
- Santos, M. Y., & Ramos, I. (2009). *Business Intelligence: Tecnologias da Informação na Gestão de Conhecimento* (2nd ed.). Lisboa: FCA.
- Skorka, A. (2017). Successful dashboard implementation in practice: How to overcome implementation barriers and ensure long-term sustainability. *International Journal of Market Research*, 59(2), 239-262.
- Turban, E., Sharda, R., Delen, D., & King, D. (2010). *Business Intelligence: A Managerial Approach* (2nd ed.). New Jersey: Pearson Education.

- Wirth, R., & Hipp, J. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proc. of 4th International Conference on the Practical Application of Knowledge Discovery and Data Mining*. (Vol. 1). London, UK: Springer-Verlag.
- Zeng, L., Xu, L., Shi, Z., Wang, M., & Wu, W. (2006). Techniques, process, and enterprise solutions of business intelligence. *In 2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 6, pp. 4722-4726. IEEE..

7. Anexos

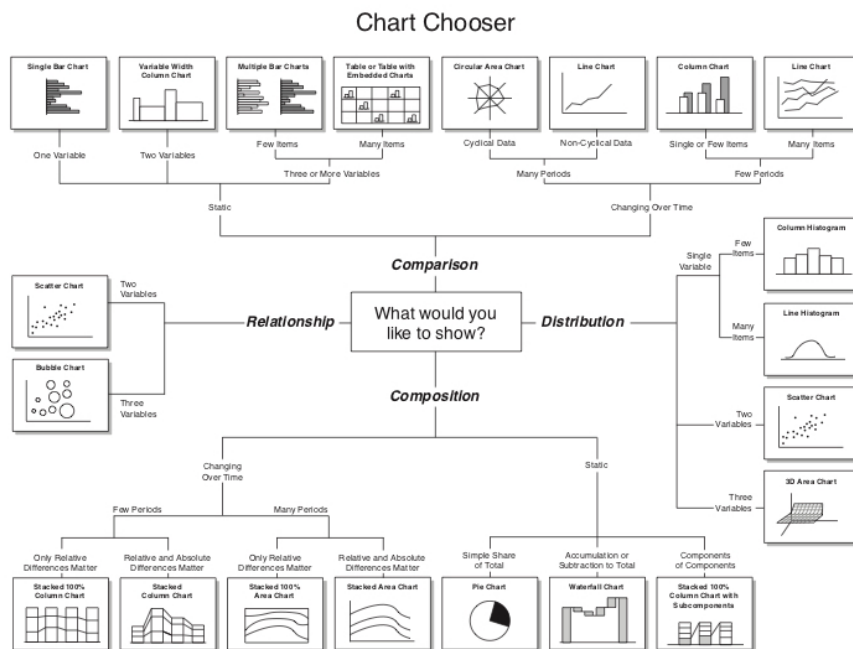


Figura 19 – Guia para a escolha do gráfico. (Fonte: Abela, 2006)



Figura 20 – Magic Quadrant for Analytics and Business Intelligence Platforms.

(Fonte: Richardson et al., 2020)

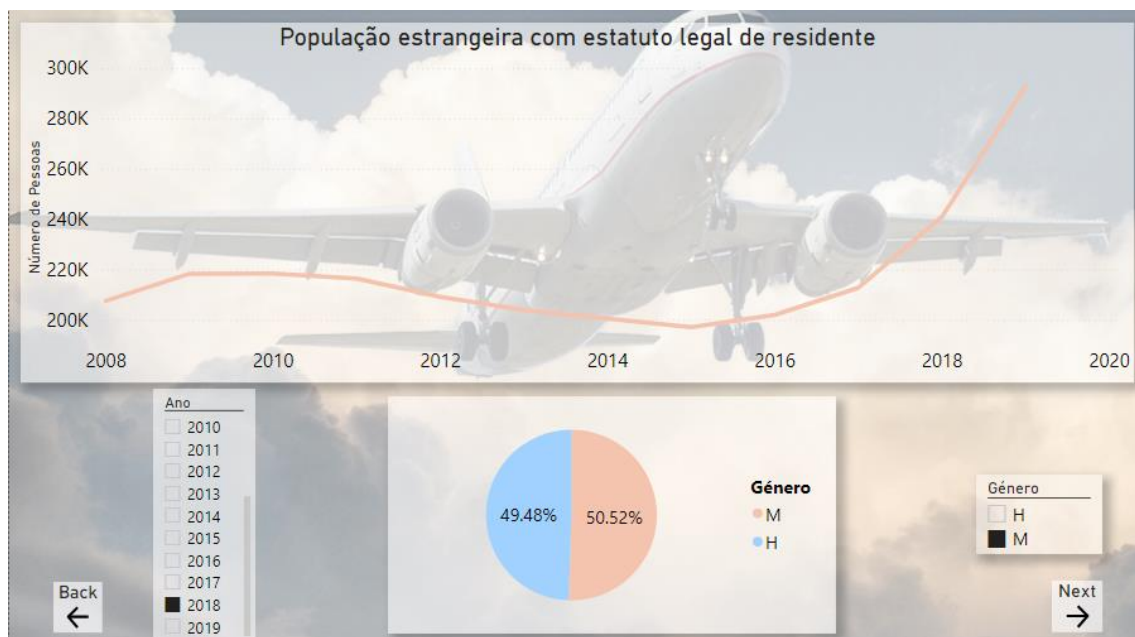


Figura 21 – Aplicação dos filtros “Género” - M e “Ano” - 2018.

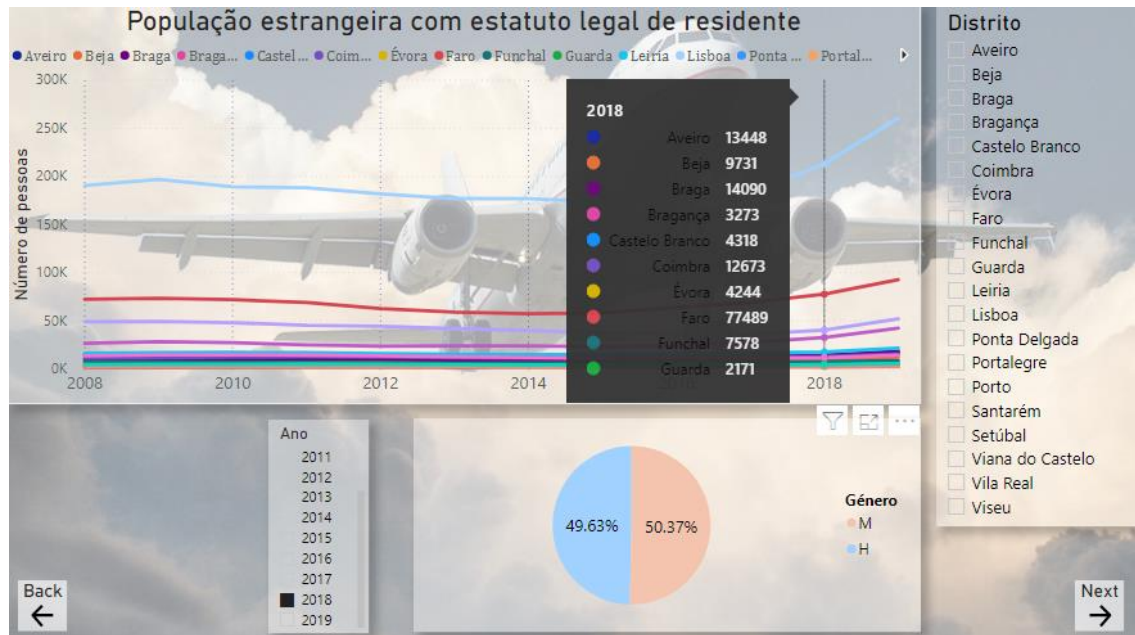


Figura 22 – População estrangeira por distrito e a respetiva percentagem por género.

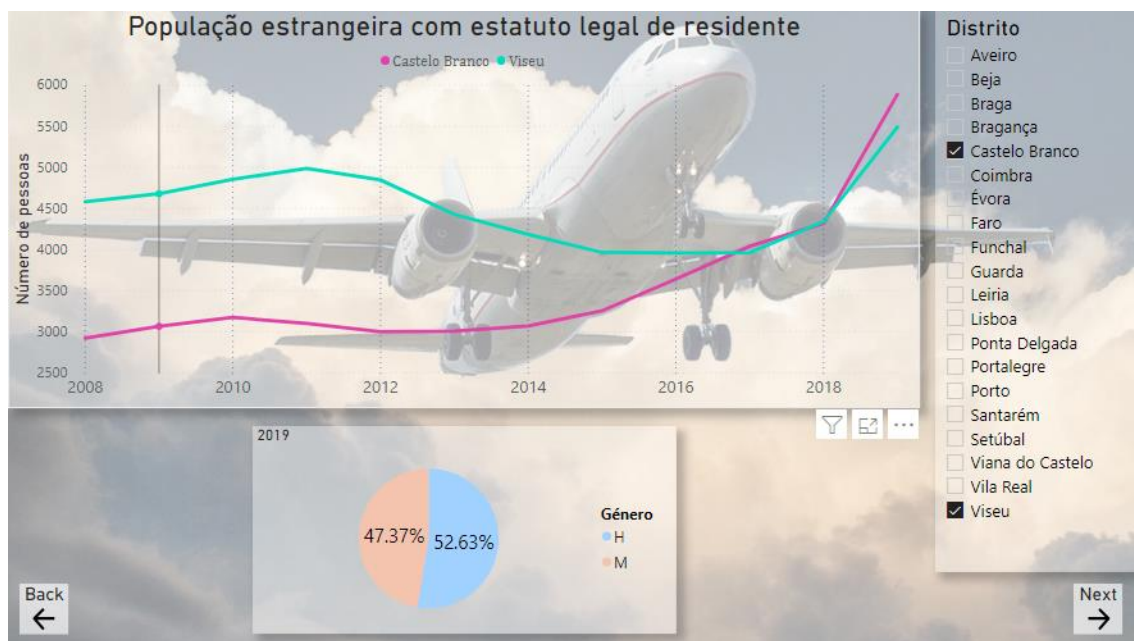


Figura 23 – Comparação da evolução da população estrangeira entre dois distritos.